Untersuchungen zur Schülerbeurteilung in Abhängigkeit von impliziten Persönlichkeitsmodellen bei Lehrern

DISSERTATION

zur

Erlangung des Grades eines Doktors
der Erziehungswissenschaften
an der
Pädagogischen Hochschule Ruhr
in Dortmund

vorgelegt von

Klaus Tscherner
Dortmund

1974

Druck- und Verlagsanstalt
Werner Blasaditsch · 89 Augsburg · Theodor-Heuß-Platz 3 · Telefon (0821) 51 85 63
Printed in Germany

Untersuchungen zur Schülerbeurteilung in Abhängigkeit von impliziten Persönlichkeitsmodellen bei Lehrern

DISSERTATION

zur

Erlangung des Grades eines Doktors
der Erziehungswissenschaften

an der

Pädagogischen Hochschule Ruhr

in Dortmund

vorgelegt von
Klaus Tscherner
Dortmund

UNIV.
LEIDEN
BIBL.

1. Gutachter: Prof. Dr. W. Piel
2. Gutachter: Prof. Dr. R.G.E. Müller
Tag der mündl. Prüfung: 12.2.1974

Kurze Zusammenfassung der Arbeit

Ausgehend von den Befunden HOFERs (1969) über die Schülerpersönlichkeit im Urteil des Lehrers bearbeiten wir - unter Verwendung von 25 Eigenschaftspaaren, mit denen 36 Lehrer die Schüler ihrer jeweiligen Klassen beurteilten - die Frage, aufgrund welcher und wievieler Beurteilungsdimensionen Lehrer die Schüler ihrer Klassen gruppieren. Clusteranalysen der Lehrerurteile über die Schüler erbrachten für jeden Lehrer konsistent drei Schülergruppen mit sehr unterschiedlichen mittleren Ausprägungsgraden auf verschieden vielen Beurteilungsdimensionen. Kriterien zur interpretativen Beschreibung unterschiedlich beurteilter Schülergruppen werden erstellt.

Für alle Lehrer lassen sich statistisch bedeutsame Zusammenhänge zwischen den fünf Beurteilungsdimensionen Arbeitsverhalten, soziale Aufgeschlossenheit vs. Verschlossenheit, Begabung, Normenkonformität und Empfindsamkeit vs. Robustheit einerseits und Zeugnisnoten in verschiedenen Fächern andererseits nachweisen, wobei die genannten Beurteilungsdimensionen in Reihenfolge und Beitrag zur Varianzaufklärung der Noten von Lehrer zu Lehrer beträchtlich variieren. Zudem zeigt sich ein Trend, daß die Zusammenhänge für Jungen höher ausfallen als für Mädchen. Es läßt sich belegen, daß das implizite Persönlichkeitsmodell HOFERs repräsentativ für die Notengebung ist.

Summary:

Using HOFER's researches (1969) into the personality of pupils as judged by the teacher as a starting point, we have attempted to answer the following question: according to what and how many dimensions do teachers classify their pupils? In the experiment 36 teachers judged their pupils, choosing between 25 pairs of characteristics. Clusteranalyses of the teachers' judgements reveal in every case three groups of pupils with widely divergent average grades of expression in a varying number of dimensions. Criteria were established for an interpretative description of divergently judged pupil groups.

Statistically significant connections can be proved for the five dimensions of judgement (attitude to work, sociability, academic ability, conformity to social norms and sensitivity vs. "sturdiness") on the one hand and term grades in different subjects on the other. The dimensions of judgement show a remarkable degree of variation from teacher to teacher as far as the sequence and the degree to which a particular dimension is taken into account for grading are concerned. For boys the correlations seem to be higher than for girls. It can be stated that the implicit model of personal constructs set up by HOFER (1969) is representative for grading.

I THEORETISCHER TEIL

1. Einleitung: Die Bedeutung von Beurteilungsprozessen in der Lehrer-Schüler-Interaktion

2. Inhaltliche und methodologische Abklärungen zum Problem der Verhaltensbeurteilung in der sozialen Interaktion

2. 1. Vorbemerkungen

2. 2. Ausgangspunkte der Forschung zur sozialen Wahrnehmung

2. 3. Methoden der Verhaltensbeurteilung im Rahmen der Forschung zur sozialen Wahrnehmung

2. 4. Fehlerquellen der Beurteilung

2. 5. Reliabilität und Validität von Verhaltensbeurteilungen

2. 6. Bedingungen, welche Reliabilität und Validität beeinflussen

2. 6. 1. Die individuelle Eigenart des Beurteilers

2. 6. 2. Der Einfluß der beurteilten Eigenschaft

2. 6. 3. Der Einfluß der sozialen Beziehung zwischen Beurteiler und Beurteiltem

2. 7. Reliabilität und Validität von Lehrerurteilen, soweit sie in Schulnoten zum Ausdruck kommen

2. 8. Erörterungen zum Problem der kognitiven Komplexität

2. 9. Folgerungen und Hinweise zum Untersuchungsablauf

II EMPIRISCHER TEIL

3. Fragestellungen

4. Durchführung der Untersuchung

4. 1. Untersuchungsmaterial

4. 2. Datenerhebung

4. 3. Datenanalyse

4. 3. 1. Faktorenanalyse des Polaritätenprofils

4. 3. 2. Clusteranalytische Ermittlung von Urteilstendenzen bei den einzelnen Lehrern

4. 3. 2. 1. Klassifikation ähnlich beurteilter Schüler mit der Hierarchical Grouping Analysis nach WARD

4. 3. 2. 2. Profildarstellung der Ergebnisse der Hierarchical Grouping Analysis

4. 3. 2. 2.1. Kriterien zur Profilinterpretation

4. 3. 2. 2.2. Interpretationen einzelner Profilabbildungen

4. 3. 2. 3. Abschließende Bemerkungen zur Interpretation einzelner Profile

4. 3. 3. Multiple Korrelationsanalysen

4. 3. 3. 1. Bemerkungen zu den instrumentellen Eigenschaften der Prädiktor- und Kriteriumsvariablen

4. 3. 3. 2. Darstellung globaler Zusammenhänge zwischen den Prädiktor- und Kriteriumsvariablen in einzelnen Klassen

4. 3. 3. 3. Darstellung der schrittweisen Varianzaufklärung der Kriteriumsvariablen durch die Prädiktorvariablen in einzelnen Klassen

5. Zusammenfassung der Ergebnisse

6. Diskussion der Ergebnisse

7. Literaturverzeichnis

8. Anhang

I. THEORETISCHER TEIL

1. Einleitung:

Die Bedeutung von Beurteilungsprozessen in der Lehrer-Schüler-Interaktion

Diese Untersuchung möchte den Lehrer mit den wichtigsten Grundprinzipien der sozialen Wahrnehmung vertraut machen und Möglichkeiten aufzeigen, wie der einzelne Lehrer Rückmeldung über sein Urteilsverhalten erlangen und wie er insbesondere erkennen kann, welche Beurteilungsaspekte in seine Notengebung einfließen.

Selbst bei sehr guter Planung und Vorbereitung von Unterrichtsstunden ist der Lehrer ständig gezwungen, spontane Entscheidungen zu fällen, Entscheidungen, die aufgrund von Verhaltensanalysen der Klasse sowie einzelner Schüler getroffen werden müssen. Diese Entscheidungen basieren auf Beurteilungen, die herkömmlicherweise in Leistungs- und Persönlichkeits- bzw. Charakterbeurteilungen eingeteilt werden. Neben den Leistungsbeurteilungen nimmt der Lehrer Beurteilungen des sozialen und emotionalen Verhaltens von Schülern in der täglichen Interaktion ständig vor, auch wenn er sie nicht oder nur gelegentlich verbalisiert. Um auf unerwünschtes Schülerverhalten angemessen reagieren zu können, muß der Lehrer in der Lage sein, die Gründe unerwünschten Schülerverhaltens zu erkennen. Wir brauchen nur daran zu denken, welche Gründe vorliegen können, wenn ein Schüler beispielsweise im Unterricht nicht mitarbeitet. Er kann in der vorangegangenen Stunde überfordert worden sein, er kann übermüdet sein, er kann an Probleme in der Familie oder mit Klassenkameraden denken, um nur einige Gründe zu nennen.

Sozialpsychologen und Psychotherapeuten weisen immer wieder darauf hin, daß Gruppen, die sich aus Mitgliedern mit genauen sozialen Wahrnehmungen zusammensetzen, nicht nur leistungsfähiger, sondern auch eher in der Lage sind, potentielle Konflikte zu vermeiden. Innerhalb der Institution Schule ist der Lehrer infolge vielfältiger Belastungen und der daraus resultierenden globalen Konfrontation mit den Schülern besonders gefährdet, vereinfachte und schematisierte Urteile zu fällen. So zeigt MÜLLER (1970, S.50-60) anhand einer gründlichen Analyse des deutschen und anglo-amerikanischen Schrifttums und empirischer Untersuchungen zum Problem der Klassifizierung von Verhaltensstörungen bei Schulkindern auf, daß Lehrer in ihren Schülerurteilen vorwiegend solche Verhaltensweisen hervorheben, "die ihren eigenen Begriff von Moral, Gehorsam, Ordnung, Lerneifer und sozialer Einordnung verletzten", wohingegen auffällig wenig Verhaltensweisen genannt werden, "die emotionale und soziale Probleme des Kindes anzeigten, ohne daß sich diese in einem Störverhalten niederschlugen." (MÜLLER, 1970, S.55) Neben einer generellen Empfindlichkeit der meisten Lehrer gegenüber uneingeordneten Verhaltens, die maßgeblich für Schülerbeurteilungen ist, weist MÜLLER (a.a.O.) auf individuelle Unterschiede bei Lehrern hinsichtlich der Empfindlichkeit gegenüber Verhaltensschwierigkeiten hin, indem er eine empirische Untersuchung WICKMANN's (1949) herausstellt: "Obwohl die Lehrer in der Cleveland-Schule ziemlich homogene Klassen hatten, zeigen ihre Berichte beträchtliche individuelle Unterschiede. So werden von zwei Lehrern 100 % der Schüler der Unordnung im Klassenzimmer bezichtigt, während sieben andere Lehrer dieses Verhalten an weniger als 15 % der Kinder feststellten. Eine

Lehrerin hält 92 % der Kinder für unehrlich, während zwei andere erklären, daß sie bei keinem Kind Anzeichen für Unaufrichtigkeit festgestellt haben. - Drei Lehrer berichten über absichtlichen Ungehorsam bei über einem Drittel der Schüler, während neun Lehrer dieses Verhalten nie bei einem ihrer Kinder beobachtet haben. - Vier Lehrer erklären, daß über 2/3 ihrer Kinder Betrügereien begangen haben, aber sechs andere glauben, daß keiner ihrer Schüler jemals betrogen habe." (MÜLLER, 1970, S.57)

Neben der Empfindlichkeit vieler Lehrer gegenüber "Problem"-Kindern, deren persönliche Anliegen übersehen und Unterrichts- und Organisationsproblemen untergeordnet werden, unterliegt die Einschätzung der Schülerpersönlichkeit durch den Lehrer einer Reihe von systematisch variierenden Fehlern, die im Einzelfall schwierig zu identifieren sind. Zu diesen Fehlern gehören nach TENT (1969, S.55)

" 1. Die irrepräsentative Entnahme von Zeitstichproben aus dem Gesamtverhalten des zu beurteilenden Individuums,

2. die Kriteriumsverzerrung durch Übergewichtung von Teilmerkmalen der beurteilten Dimension und

3. die systematischen Einstellungsfehler, wie der bekannte Halo-Effekt, oder die Tendenz zur Skalenmitte oder zur Milde bzw. Härte der Beurteilung (siehe auch O.ORLIK 1961, 1967; LIENERT & Anna Dorothea HOPP 1964)."

HOFER (1969) wies mit Hilfe der Faktorenanalyse und Techniken der multidimensionalen Skalierung nach, daß Lehrer ihre Schüler nach einem impliziten

Persönlichkeitsmodell beurteilen. Durch die folgenden fünf Eigenschaftsdimensionen läßt sich dieses implizite Persönlichkeitsmodell beschreiben:

I. Diszipliniertheit vs. Undiszipliniertheit;
II. Geistige Regsamkeit;
III. Bescheidene Zurückhaltung vs. Geltungsstreben;
IV. Soziale Aufgeschlossenheit vs. Verschlossenheit;
V. Empfindsamkeit vs. Robustheit.

Diese subjektive Beurteilungsstruktur gestattet es den Lehrern, das Leistungs- und Persönlichkeitsbild ihrer Schüler mit relativ wenigen Informationen zu charakterisieren.

In der psychologischen Diagnostik lassen sich die impliziten Persönlichkeitstheorien, wie z.B. das Modell von HOFER, neben der allgemeinen Beurteilungstendenz, dem Halo-Effekt und den Kontrast-Ähnlichkeitsfehlern in die Reihe der Beurteilungsfehler einordnen. Die Urteile über den Mitmenschen werden durch das jeweilige Persönlichkeitsmodell der Beurteiler beeinflußt, d.h. diese haben eine bestimmte Vorstellung davon, wie verschiedene Merkmale "in der Natur" korreliert sind ("Wer lügt, stiehlt auch"). Implizite Persönlichkeitsmodelle bzw. -theorien geben also Aufschluß über die Art und Weise, in der Menschen ihre Erfahrungen mit Mitmenschen zu bestimmten mehr oder weniger konsistenten Vorstellungen über diese verarbeiten (vgl. BIERI et al. 1966, BRUNER, SHAPIRO und TAGUIRI 1958, KAMINSKI 1959, KELLY 1955, SARBIN, TAFT und BALLEY 1960).

Daß den Vorstellungen von Beurteilern über die "normale" Persönlichkeit keine differentialdiagnostische Bedeutung zukommt, wird in einem interessanten Experiment von BECKER (1960) deutlich. Der genannte Autor ließ dieselben Kinder auf denselben Skalen einmal von den Eltern, einmal von den Kindergärtnerinnen beurteilen. Zwei getrennte R-Analysen ergaben praktisch den gleichen Merkmalsraum; daraus darf jedoch nicht der Schluß gezogen werden, daß die beiden Rater-Gruppen gleich geurteilt hätten. Verglich man nämlich die Faktorenscores der Kinder der beiden Analysen, so ergaben sich zwischen ihnen Korrelationen um null. Die Faktorenstruktur spiegelte also das gleiche semantische System der Beurteiler und deren Vorstellungen über die "normale" Persönlichkeit, die Ratings selbst aber hatten keine differentialdiagnostische Bedeutung.

In dieser Untersuchung soll jedoch nicht die Frage bearbeitet werden, inwieweit Beurteilungen von Lehrern über ihre Schüler "richtig" sind, d.h. Tatsachen abbilden. Da man für viele Eigenschaften nicht auf objektive Kriterien zurückgreifen kann, läßt sich diese Frage schwer beantworten. Anhaltspunkte, daß die implizit vermuteten Eigenschaftszusammenhänge, die bei Lehrern hinsichtlich der Schülerbeurteilungen zum Tragen kommen, gerade für den Persönlichkeitsbereich die "wahren" Zusammenhänge falsch widerspiegeln, ergeben sich aus einer Untersuchung von BARTRAM, ROLLETT und SUHRE (1971). Demnach sind Lehrer, wenn Testergebnisse zum Vergleich herangezogen werden, in der Lage, Intelligenz und Arbeitsverhalten ihrer Schüler eher "richtig"

einzustufen als Persönlichkeitsmerkmale wie beispielsweise Neurotizismus, Extraversion und Ängstlichkeit. Vielmehr werden Neurotizismus und Ängstlichkeit bei Jungen im Zusammenhang mit mangelndem, bei Mädchen dagegen mit diszipliniertem Arbeitsverhalten gesehen. Mädchen mit hoher Ausprägung auf der Dimension Extraversion werden von Lehrern als geltungsbedürftig, unbegabt und unsympatisch abqualifiziert.

Derartige Befunde sollten Pädagogen und Lehrer nachdenklich stimmen, zumal dann, wenn, wie STEINKAMP (1965) feststellen konnte, 78 % der Hamburger Lehrer der Meinung sind, daß bei Empfehlungen zum Besuch einer weiterführenden Schule neben der Leistungsbewertung auch die A r b e i t s h a l t u n g und der C h a r a k t e r (!) der Schüler berücksichtigt werden müsse und wenn man die geringe Interkorrelation zwischen Persönlichkeitsvariablen einerseits und zwischen Leistungs- und Persönlichkeitsmerkmalen andererseits in Rechnung stellt. Solche Untersuchungen machen deutlich, daß in Lehrergutachten über Schüler in erster Linie die "impliziten Theorien" der Lehrer zum Ausdruck kommen.

Als ein weiterer wesentlicher Faktor der interpersonellen Wahrnehmung wird in zunehmendem Maße die Differenziertheit der Personenwahrnehmung bzw. -beurteilung erkannt (vgl. u.a. BIERI et al. zur Bedeutung der Differenziertheit im Rahmen der klinischen Urteilsbildung). Unter Differenziertheit wird dabei die Anzahl unabhängiger Dimensionen verstanden, über die ein Individuum bei der Kategorisierung aufgenommener Reize verfügt. In dieser Arbeit

versuchen wir, Aufschlüsse über die Urteilsdifferenziertheit von Lehrern im Rahmen der von HOFER (1969) gefundenen Eigenschaftsdimensionen zur aktuellen Schülerbeurteilung zu erlangen. Mit Hilfe von Clusteranalysen wird zu überprüfen sein, wie Lehrer im einzelnen ihr implizites Persönlichkeitsmodell realisieren, d.h. aufgrund welcher Eigenschaftsdimensionen welche Lehrer zu unterschiedlichen Schülerbeurteilungen kommen und welche Bewertungsaspekte in die Notengebung einfließen. Auf die speziellen Fragestellungen wird in Kap. 3 näher eingegangen werden. Vorerst jedoch werden wir uns einigen methodischen und inhaltlichen Erörterungen zum Forschungsgegenstand der sozialen Wahrnehmung zuwenden.

2. Inhaltliche und methodologische Abklärungen zum Problem der Verhaltensbeurteilung in der sozialen Interaktion

2.1. Vorbemerkungen

Die Beurteilung der Mitmenschen hat im wissenschaftlichen und vorwissenschaftlichen Sinn eine Bedeutung. Außerhalb der wissenschaftlichen Psychologie existieren gewisse vorwissenschaftliche Modellvorstellungen über das "Bild vom Anderen" (KAMINSKI, 1959). Eine einfache Annahme besagt, daß es individuelle Unterschiede hinsichtlich der Eigenschaft der Persönlichkeitsbeurteilung (also gute oder schlechte Menschenkenner) gibt. Eine Vorstellung ist beispielsweise, daß Frauen bessere Menschenkenner sind als Männer. Die Frage, was einen guten vom

schlechten Menschenkenner unterscheidet, löst eine Reihe von Fragen aus. In der Psychologie gibt es ein weit ausgebautes Forschungsgebiet, das sich mit diesen Fragen befaßt.

2.2. Ausgangspunkte der Forschung zur sozialen Wahrnehmung

In Deutschland wurden Untersuchungen vor allem unter dem Gesichtspunkt der sog. Ausdruckspsychologie durchgeführt. Danach ist das "Bild vom Anderen" Ergebnis eines speziellen Vorgangs des Verständnisses von äußeren Erscheinungen, wozu die gängige Vorstellung gehört: "Die Seele ist immer innen drin und der Geist knapp unter den Haaren!". Um die inneren Eigenarten bestimmen zu können, sucht man nach bestimmten Augenblicken, in denen sie durch äußere Bewegungen charakterisiert sind. In der vorwissenschaftlichen Psychologie gibt es darüber konkrete Vorstellungen: "Die Seele äußert sich z.B. im Blick" ("beseelter Blick"). Direktes Ansehen gilt als Zeichen der Ehrlichkeit (Folge: Jedes unehrliche Kind oder jedes Kind, das Schwierigkeiten mit seinen Eltern hat, lernt es kurz über lang meistens sehr schnell, unwahre Angaben mit offenem und freiem Blick zu machen).

Tests, die von Psychologen zur Erfassung von Persönlichkeitseigenarten konstruiert wurden, versuchte man, am Kriterium der Beurteilung der Mitmenschen zu validieren. Das, was hier beurteilt wird, stammt im Grunde aus der Alltagspsychologie.

Weiterhin interessiert die Beurteilung des Mitmenschen für praktische Zwecke in Behörden und Firmen, wenn Vor-

versuchen wir, Aufschlüsse über die Urteilsdifferenziertheit von Lehrern im Rahmen der von HOFER (1969) gefundenen Eigenschaftsdimensionen zur aktuellen Schülerbeurteilung zu erlangen. Mit Hilfe von Clusteranalysen wird zu überprüfen sein, wie Lehrer im einzelnen ihr implizites Persönlichkeitsmodell realisieren, d.h. aufgrund welcher Eigenschaftsdimensionen welche Lehrer zu unterschiedlichen Schülerbeurteilungen kommen und welche Bewertungsaspekte in die Notengebung einfließen. Auf die speziellen Fragestellungen wird in Kap. 3 näher eingegangen werden. Vorerst jedoch werden wir uns einigen methodischen und inhaltlichen Erörterungen zum Forschungsgegenstand der sozialen Wahrnehmung zuwenden.

2. Inhaltliche und methodologische Abklärungen zum Problem der Verhaltensbeurteilung in der sozialen Interaktion

2.1. Vorbemerkungen

Die Beurteilung der Mitmenschen hat im wissenschaftlichen und vorwissenschaftlichen Sinn eine Bedeutung. Außerhalb der wissenschaftlichen Psychologie existieren gewisse vorwissenschaftliche Modellvorstellungen über das "Bild vom Anderen" (KAMINSKI, 1959). Eine einfache Annahme besagt, daß es individuelle Unterschiede hinsichtlich der Eigenschaft der Persönlichkeitsbeurteilung (also gute oder schlechte Menschenkenner) gibt. Eine Vorstellung ist beispielsweise, daß Frauen bessere Menschenkenner sind als Männer. Die Frage, was einen guten vom

schlechten Menschenkenner unterscheidet, löst eine Reihe von Fragen aus. In der Psychologie gibt es ein weit ausgebautes Forschungsgebiet, das sich mit diesen Fragen befaßt.

2.2. Ausgangspunkte der Forschung zur sozialen Wahrnehmung

In Deutschland wurden Untersuchungen vor allem unter dem Gesichtspunkt der sog. Ausdruckspsychologie durchgeführt. Danach ist das "Bild vom Anderen" Ergebnis eines speziellen Vorgangs des Verständnisses von äußeren Erscheinungen, wozu die gängige Vorstellung gehört: "Die Seele ist immer innen drin und der Geist knapp unter den Haaren!". Um die inneren Eigenarten bestimmen zu können, sucht man nach bestimmten Augenblicken, in denen sie durch äußere Bewegungen charakterisiert sind. In der vorwissenschaftlichen Psychologie gibt es darüber konkrete Vorstellungen: "Die Seele äußert sich z.B. im Blick" ("beseelter Blick"). Direktes Ansehen gilt als Zeichen der Ehrlichkeit (Folge: Jedes unehrliche Kind oder jedes Kind, das Schwierigkeiten mit seinen Eltern hat, lernt es kurz über lang meistens sehr schnell, unwahre Angaben mit offenem und freiem Blick zu machen).

Tests, die von Psychologen zur Erfassung von Persönlichkeitseigenarten konstruiert wurden, versuchte man, am Kriterium der Beurteilung der Mitmenschen zu validieren. Das, was hier beurteilt wird, stammt im Grunde aus der Alltagspsychologie.

Weiterhin interessiert die Beurteilung des Mitmenschen für praktische Zwecke in Behörden und Firmen, wenn Vor-

gesetzte Urteile über Untergebene abgeben müssen. Im Jahre 1906 wurde von HEYMAS und WIERSMA eine Untersuchung durchgeführt, in der Unterschiede zwischen Männern und Frauen interessierten. Man ließ gute Bekannte Urteile abgeben und wollte aufgrund dieser Urteile die Unterschiede gefunden haben.

Das Gebiet der sozialen Wahrnehmung hat also sehr viele Ausgangspunkte. Daher ist das Forschungsgebiet sehr zerstreut. Seit dem zweiten Weltkrieg läßt sich aber eine einheitliche Grundkonzeption erkennen. Einige methodische Forderungen wurden ausgearbeitet. Man betrachtet die Beurteilung der Mitmenschen als abhängig von

- der Eigenart des Beobachteten,
- der Eigenart des Beobachters und
- von der Eigenart der Information, die der Beobachter über den Beobachteten bekommen hat.

Die Frage nach der Richtigkeit der Urteile wird abgelöst von der Frage: Wie spielt sich der Prozeß der Urteilsbildung ab?

2.3. Methoden der Verhaltensbeurteilung im Rahmen des Forschungsgegenstandes soziale Wahrnehmung

Die Methoden der Verhaltensbeurteilung können nach folgenden Gesichtspunkten gegliedert werden:

(1) Nach Art der Beurteiler:

Hier kann die Beurteilung durch geschulte Experten erfolgen (z.B. Psychologen, Psychiater); Vorgesetzte können Untergebene beurteilen; beim sog. Peer Rating beurteilen Mitglieder einer Gruppe Gleichgestellte, z.B. Lehrer Lehrer, Studenten Studenten usw.

(2) Nach den Beobachtungssituationen:

Einerseits kann das typische Verhalten eines Pb in einer näher spezifizierten Situation geschildert werden (z.B. in Leistungssituationen, in Gefahr, nach Frustration usw.), andererseits können nach der sog. "time sampling method" (vgl. OLSON et al. 193[illegible] Zufallsstichproben gewonnen werden, von denen auf das Gesamtverhalten geschlossen werden kann. Mit der "critical incident technique" nach FLANAGAN (1954) sammelt man Berichte über alle Fälle, in denen ein bestimmtes Verhalten auftrat und analysiert die auslösenden Bedingungen. Außerdem können Pbn in einer konstruierten Realsituation beobachtet werden.

(3) Nach den Beurteilungsmerkmalen:

Die Dimensionen der Beurteilung können freigestellt oder auch sehr global sein. Besser ist es, wenn es sich um einzelne, genau definierte Merkmale handelt. Ein weiterer Schritt zur Standardisierung besteht darin, auch die verschiedenen Ausprägungsgrade eines Merkmals - möglichst operational - zu definieren.

(4) Nach den Methoden der Datenerhebung und -analyse

Mit den reinen Beobachtungsmethoden (Verhaltenslisten, Fragebögen, Verhaltensprotokollen, Beobachtungsschemata) wird Verhalten beschrieben, nicht beurteilt. Die sog. relativen Beurteilungsmethoden (Rangreihen, "master scales", "man-to-man scale", Paarvergleich, Soziometrie) vergleichen jeden Pb mit den anderen Pbn einer Stichprobe. Mit den sog. absoluten Beurteilungsmethoden ("check lists", "forced-choice-ratings", "portrait matching", "guess-who-technique", Q-Sort, rating scales, Prozentrangskalen) soll ein Pb unabhängig von seiner Stellung innerhalb einer Gruppe beurteilt werden.

Bei der völlig ungebundenen Beurteilung sind dem Beurteiler die Beurteilungsdimensionen freigestellt, daher ist ein Vergleich zwischen verschiedenen Beurteilern schwierig. Ein weiterer Nachteil ist, daß der jeweils wichtigste Gesichtspunkt vom Beurteiler abhängig ist. Weiterhin spielt die Reihenfolge, in der Beschreibungsgesichtspunkte gebracht werden, eine Rolle, was in einer berühmt gewordenen Untersuchung von ASCH (1946) nachgewiesen wurde. ASCH setzte an den Anfang oder das Ende einer Reihenfolge von Wörtern wie "intelligent", "fleißig" usw. jeweils die Worte "warm" oder "kalt". Der Eindruck, der beim Hören der Wortreihe entstand, war abhängig davon, ob "warm" oder "kalt" am Anfang oder am Ende der Liste stand.

2.4. Fehlerquellen der Beurteilung

Wenn unabhängige Rater dieselben Pbn hinsichtlich eines Merkmals beurteilen, so können die Mittelwerte der Urteile pro Rater differieren, z.B. dadurch, daß einzelne Beurteiler generell zu streng oder zu milde urteilen. In der Tat scheint dieser Fehler innerhalb von Ratern einigermaßen konstant zu sein (vergl. DUBIN et al., 1954). Man könnte ihn dann durch einfache Mittelwertstransformation beseitigen. GUILFORD (1954) spricht bei Milde und Strenge von einem positiven bzw. negativen "Error of Liniency", CRONBACH (1958) vom "Generosity Error". Die Tendenz zur günstigen Beurteilung findet man am häufigsten bei "Charakterbeurteilungen", die zur schlechteren Beurteilung beim Bewerten von Leistungen. Die Tendenz zur mittleren Beurteilung findet man besonders bei unsicheren Beurteilern, die auf diese Weise keine großen Fehler begehen können.

Der Name "Halo-Effekt" (THORNDIKE, 1920) ist eine Bezeichnung für die Tendenz, einzelne Merkmale aufgrund eines allgemeinen Gesamturteils einzustufen, z.B. werden einem allgemein sympathischen Menschen alle wünschenswerten Eigenschaften zugeschrieben. Varianzanalytisch ausgedrückt handelt es sich um eine Wechselwirkung zwischen Rater und Pb: Der Rater über- oder unterschätzt den Pb auf allen Dimensionen. Der Halo-Effekt erhöht die Korrelationen zwischen verschiedenen Beurteilungsskalen. Man kann ihn teilweise dadurch reduzieren, daß man erst alle Pbn auf einer Skala beurteilen läßt, dann auf der nächsten usw., ohne daß zurückgeblättert werden darf. Es gibt wenig empirische Untersuchungen über den Halo-Effekt, die aber alle zu klaren Bestätigungen führen (vergl. u.a. BRUNSWIK 1945, EARLE 1926, JURGENSON 1950, KNIGHT 1923).

Kontrast-/Ähnlichkeitsfehler liegen vor, wenn die Urteile abhängig vom impliziten Selbsturteil gegeben werden. Rater, die sich für sehr intelligent oder ordentlich halten, könnten u. U. geneigt sein, andere in dieser Hinsicht zu unterschätzen. Andererseits gibt es eine Tendenz, andere für uns selbst ähnlich zu halten, z.B. neigen Pbn, die auf der F - Skala hoch liegen, dazu, auch andere als autoritär einzuschätzen (SCODEL und FREEDMAN, 1956).

Die impliziten Persönlichkeitstheorien wurden eingangs schon in die Reihe der Beurteilungsfehler eingeordnet. Beurteiler und Beurteilter können in ihrer Auffassung, welche Eigenschaften zusammengehören und welche nicht, übereinstimmen ("Rothaarige sind leidenschaftlich"). Die Beurteilung und das Selbsturteil

stimmen nur aufgrund der gleichen impliziten Persönlichkeitstheorie überein. Hieraus schließt man, daß der Beurteiler die individuelle Eigenart besäße, richtige Vorhersagen zu treffen. Überprüft man aber, welche Menschen er gut beurteilen kann, so stellt sich heraus, daß es nur solche sind, die über die gleiche Persönlichkeitstheorie verfügen.

Übereinstimmungen zwischen verschiedenen Beurteilern können u.U. als Wirkung gemeinsamer Stereotype interpretiert werden (RICE 1927, LITTERER 1933, CUNRNEE 1934, CENTERS 1951, CHAREN 1955, MASON 1957). Diese Stereotype bleiben auch dann erhalten, wenn zur Beurteilung nur wenige Informationen zur Verfügung stehen, die mit dem beurteilten Merkmal wenig zu tun haben, wie etwa der Klang der Stimme für die Beurteilung des Berufes (vergl. FAY und MIDDLETON, 1929) oder etwa das Tragen einer Brille (vergl. u.a. THORNTON 1943 und 1944, STROBL 1956).

Merkmale, die unmittelbar hintereinander eingestuft werden, korrelieren höher als solche mit größerem zeitlichen Abstand (STOCKFORD und BISSELL, 1949). Aus diesem Grunde ist es empfehlenswert, heterogene Merkmale hintereinander beurteilen zu lassen oder generell die Zeitabstände zu erhöhen.

2.5. Reliabilität und Validität von Persönlichkeits- und Verhaltensbeurteilungen

Unter Reliabilität von Verhaltenseinstufungen kann verstanden werden:

- Das Ausmaß, zu welchem Beurteilungen der gleichen Beurteiler zu verschiedenen Zeitpunkten übereinstimmen.

- Das Ausmaß, zu welchem die Aussagen verschiedener Beurteiler über die gleichen Personen übereinstimmen.

Zur Reliabilitätsschätzung benutzt man Retest-Korrelationen und die Inter-Rater-Übereinstimmung (vergl. GUILFORD 1954, S. 395 f.). Im ersteren Fall liegen die Koeffizienten gewöhnlich bei .70 bis .80, im letzteren bei .50 bis .60. Die Inter-Rater-Korrelation wird gesenkt, wenn es Interaktionen zwischen Ratern und Pbn gibt; ebenfalls dann, wenn den Ratern unterschiedlich viele Informationen zur Verfügung stehen. Sie wird erhöht durch gemeinsame Persönlichkeitsstereotype.

Die Stabilität wird gesenkt durch Merkmalsschwankungen beim Pb, durch Schwankungen in der impliziten Definition des Merkmals und durch Änderung der Maßstäbe beim Rater. Sie wird erhöht durch Erinnerungseffekte.

Die Reliabilität von gemittelten Urteilen (aufgrund mehrerer Rater) wächst allgemein mit der Anzahl der Rater. Bei mehr als fünf bis sieben Ratern wird der Reliabilitätszuwachs allerdings unbedeutend.

Die Frage nach der Fähigkeit, andere valide beurteilen zu können (Validitätsaspekt), hängt eng mit der Wahl des Kriteriums, an dem man die Fähigkeit mißt, zusammen. Als Kriterium wurde gelegentlich die Zu-

stimmung des Beurteilten genommen. Es läßt sich aber leicht nachweisen, daß dieses Kriterium völlig unbrauchbar ist. Studenten mußten eine Handschriftenprobe abgeben, die sie nach zwei Wochen mit einem Gutachten versehen zurückerhielten. Das Gutachten war aus einem Buch von KLAGES entnommen worden. Jeder Student erhielt dasselbe Gutachten, und es stellte sich heraus, daß fast alle Studenten alles als richtig anstrichen, nachdem man die Studenten aufgefordert hatte, das, was sie für sich als zutreffend erachteten, anzustreichen. Ähnlich dürften Aussagen von der Art: "Sie befinden sich in einer Situation, in der Sie Ihre eigentlichen Fähigkeiten nicht voll entfalten können", für fast alle Leute zutreffen. Nicht viel besser steht es mit der Zustimmung von Fachpsychologen, die der gleichen Schule bzw. dem gleichen Institut entstammen, so daß sich mehr die Übereinstimmung zwischen Spezialisten gleicher Schule als die Richtigkeit der Beurteilung zeigt.

Wenngleich Testergebnissen als Kriterium eher zuzustimmen sind, sprechen doch einige Bedenken dagegen (MERZ 1962, S. 34). Wenn beispielsweise die Intelligenzbeurteilungen von Lehrern mit Intelligenztests verglichen werden, ist zu hinterfragen, ob Lehrer das gleiche unter Intelligenz verstehen wie der Test. Die Eigenschaften, die ein Test erfaßt und die, die beurteilt werden, können u. U. verschieden sein. Weiterhin sind Testergebnisse analytischer Natur, während die Urteile des Menschen komplexer sind. Nach TAFT (1955) korrelieren Intelligenztestergebnisse mit der Beurteilungsfähigkeit von .30 bis .40. LUFT (1950) fand keine signifikanten Unterschiede zwischen Psychiatern, Sozialfürsorgern, klinischen Psychologen und Naturwissenschaftlern, die den Persönlichkeitsfragebogen eines (!) Patienten voraussagen sollten, nachdem

dessen Therapeut ihn geschildert hatte. Das gleiche ergab sich, wenn Rater der obigen Gruppen die Abschriften je eines klinischen Interviews mit zwei Patienten als Information bekamen. Im Mittel schnitten jedoch die Naturwissenschaftler am besten ab. TAFT (1950) fand eine Überlegenheit von experimentell tätigen Wissenschaftlern (auch von Psychologen) gegenüber anderen Wissenschaftlern. Allgemein scheinen Sachbezogenheit und soziale Unabhängigkeit die genaue Vorhersage von Testdaten eines Probanden zu begünstigen. Die Dauer der Beobachtung muß die Richtigkeit der Beurteilung nicht notwendig beeinflussen. Offiziersanwärter schätzen einander schon nach fünf Tagen Bekanntschaft genauso gut ein wie zu jedem späteren Zeitpunkt (HOLLANDER, 1955).

Zur Überprüfung der Frage, ob es eine unterschiedliche Beurteilungsfähigkeit gibt und damit die Frage nach der Validität überhaupt sinnvoll zu stellen ist, hat man sich weiterhin der sog. "Empathy Tests" und der Deutung des Gesichtsausdrucks auf Photos bedient. Im Falle der "Empathy Tests" muß der Rater voraussagen, wie ein Pb die Items eines Persönlichkeitsfragebogens beantworten wird. Das Kriterium für die Richtigkeit ist der Grad der Übereinstimmung zwischen der Vorhersage des Beurteilers und der tatsächlichen Beantwortung. Für dieses Verfahren spricht, daß es objektiv und einfach ist und dem operationalem Denken entgegen kommt, wenngleich das Selbstbild und nicht die Persönlichkeit des Pb "erkannt" wird. Jedoch braucht der Grad der Übereinstimmung nichts zu besagen. Ein Beurteiler kann beispielsweise deswegen zu diskrepanten Ergebnissen kommen, weil er einen anderen subjektiven Durchschnitt als der Beurteiler hat. Auf die Frage: "Haben Sie häufig Streit mit Ihrem

Vater?" kann der Beurteilte auf einer Skala z.B. in der Mitte ankreuzen, der Beurteiler aber würde weiter draußen, bei "häufig" ankreuzen, obwohl auch er die gleiche "Menge an Streit" meint. Die Bezugspunkte von Beurteiler und Beurteiltem können also verschieden sein.

Da sich die Untersuchungen auf die Beurteilung mehrerer Menschen bezieht, hat ein Beurteiler, dem es gelingt, den Durchschnitt richtig einzuschätzen, sehr viele richtige Urteile. Man könnte folgendermaßen interpretieren: Der Beurteiler hat eine geringe Diskrepanz, also ist er ein guter Menschenkenner. In Wirklichkeit kann er aber nur den Durchschnitt gut abschätzen.

Der zuletzt genannte Fehler spielt häufig in der sog. Bewährungskontrolle eine Rolle. Eine Arbeit von RASCH (1957), in der das Urteil von Graphologen mit dem zwei Jahre später erfolgten Urteil der betrieblichen Vorgesetzten verglichen wurde, erbrachte eine Übereinstimmung von 70 % zwischen den Urteilen der Graphologen und der Vorgesetzten. Da in dieser Untersuchung bezüglich der Vorgesetztenurteile eine Skalenabweichung von 1 noch als übereinstimmend mit der Aussage des Graphologen gilt, fallen rein zufällig 67 bis 68 % aller Urteile zusammen. Eine ähnliche Situation ergibt sich, wenn man bei der Vorhersage von Schulnoten die 3 wählt und eine Abweichung von 1 toleriert. Man würde sich nur noch in den wenigen Fällen von 1 und 6 irren. Auch die Selbstbeurteilung enthält eine Abschätzung des Durchschnitts. Haben Beurteiler und Beurteilter das gleiche allgemeine Bild vom Anderen, so ergibt sich eine rein zufällige Übereinstimmung. Kor-

relationskoeffizienten, die zwischen vielen Skalen errechnet werden, sind deshalb kein Maß für die Richtigkeit der Menschenkenntnis, sondern lediglich der Ausdruck für die richtige Einschätzung des Durchschnitts, von dem die Beurteilung ausging.

Im Falle der Deutung des Gesichtsausdrucks auf Photos sei an eine Untersuchung NOSSBERGERS (1959) erinnert. NOSSBERGER fotografierte 40 Knaben im Alter von 11 Jahren, während sie, um die Situation konstant zu halten, ein Intelligenzproblem lösten. Außerdem wurden sie mit einer Kurzform des HAWIK getestet. Die Portraits wurden von je 100 männlichen und weiblichen Beurteilern (davon 25 % Lehrer, Erzieher und Psychologen) hinsichtlich der Intelligenz auf einer Fünf-Punkte-Skala beurteilt, und zwar zunächst bei einer Expositionszeit von fünf Sekunden, dann bei beliebig langer Zeit. Die Korrelation zwischen Urteil und Intelligenztest beträgt im ersten Fall .30, im zweiten Fall .40. 27 Knaben wurden 30 Beurteilern vorgeführt, wobei aber nur das Gesicht sichtbar war. Hier beträgt die Korrelation zwischen Urteil und Test .37. LEITNER (1956) ließ neun Personen mit sehr verschiedener Intelligenz, aber ungefähr gleichem Bildungsgrad, drei Sprechproben abgeben: a) sinnfreie Wörter; b) einen Stifter-Text; c) eine freie Nacherzählung. 75 Studenten beurteilten die Intelligenz der Sprecher auf einer Fünf-Punkte-Skala. Die Korrelationen zwischen Urteil und Testwert betragen bei sinnfreien Wörtern .71, beim Stifter-Text .60 und bei freier Nacherzählung .79. Von dieser letzten Untersuchung abgesehen kann gesagt werden, daß für die Beurteilungen Validitätskoeffizienten zwischen etwa 0.00 und 0.50 gefunden werden. Koeffizienten um .30 können als bezeichnend gelten.

Der durchschnittliche Anteil der Beurteilervarianz bei Ratings ist noch unbekannt. Wenn mehrere Rater mehrere Pbn auf mehreren Skalen beurteilen, kommen als Varianzquelle infrage: Rater, Pbn, Skalen und die Wechselwirkungseffekte. Die Varianz zwischen Skalen und die Wechselwirkungen mit dem Faktor Skala lassen sich nicht so ohne weiteres berechnen, da es sich um unterschiedliche Metriken handelt. Im Idealfall dürfte es weder Inter-Rater-Varianz noch Wechselwirkungen zwischen Ratern und den beiden anderen Faktoren (Pbn und Skalen) geben; die gesamte Varianz müßte auf Unterschiede zwischen Pbn und zwischen Skalen sowie auf Wechselwirkungen zwischen Pbn und Skalen zurückgehen. Das dürfte jedoch kaum jemals der Fall sein. Das übliche Vorgehen, die Urteile über die Rater zu mitteln, verschleiert diesen Sachverhalt allenfalls, ohne ihn zu ändern. Bei Faktorenanalysen sollte man dreidimensional (nach Skalen, Pbn und Ratern) analysieren.

Zusammenfassend läßt sich herausstellen, daß die Validität von Beurteilungen mangels Kriterien schwer abzuschätzen ist. Außerdem entsteht ein Zirkelschluß, wenn als Kriterium ein Test verwendet wird, der an Beobachtungen validiert wurde. Verhaltensbeurteilungen sollten nur dann als diagnostisches Instrument benutzt werden, wenn es keine bessere Methode gibt, und in diesem Fall möglichst systematische, standardisierte Beschreibungsverfahren.

2.6. Bedingungen, welche Reliabilität und Validität beeinflussen

2.6.1. Die individuelle Eigenart des Beurteilers

In Kap. 2. 4. wurden bereits die sog. Beurteilungsfehler als allgemeine Eigentümlichkeiten der Persönlichkeitsbeurteilung dargelegt.

Einer verbreiteten Ansicht zufolge gibt es individuelle Unterschiede hinsichtlich der "Menschenkenntnis", des "Ausdrucksverständnisses" bzw. der "Einfühlungsfähigkeit". Zahlreiche Untersuchungen bemühen sich um die Frage, durch welche Eigenschaften der gute Menschenkenner ausgezeichnet sei. Großen Einfluß auf diese Untersuchungen hatte ALLPORT (1937, S. 526 f. der acht Eigenschaften nennt: Die Erfahrung des Beurteilers, die Ähnlichkeit des Beurteilers mit dem Beurteilten, die Einsicht des Beurteilers in seine eigenen Schwächen, eine ästhetische Haltung, das Ausmaß der kognitiven Differenziertheit, Intelligenz und soziale Distanz und kühle Introvertiertheit des Beurteilers.

Eine große Anzahl empirischer Arbeiten belegt, daß der "erfahrene Mensch" nicht besser als ein Mensch mit vergleichsweise weniger Erfahrung urteilt (vergl. u.a. ARLITT 1926, LÄMMERMANN 1927, CAREY, BERG und VAN DUSEN 1951, DYMOND 1953/54, REMMERS 1934, BUSS, WIENER, DURKEE und BAER 1955, KESSEN 1957, STONE und KAMIYA 1957, HELPER 1958).
SNOW (1924) ließ erfahrene Geschäftsleute Bewerber für einen bestimmten Posten beurteilen. Die Beurteilungen der Geschäftsleute waren nicht besser als die von Laien. KESSEN (1957) fand, daß Personen mit eigenen Kindern oder Kinderpsychologen Kinder nicht besser beurteilen als Anfängerstudenten. Psychotherapeuten können die

Reaktionen ihrer eigenen Patienten nicht voraussagen (DYMOND 1953/54), Eltern nicht die ihrer eigenen Kinder (HELPER 1958). Lehrer erreichen bei der Beurteilung des Schulerfolges ihrer Schüler nur niedrige Validitätskoeffizienten, worüber in einem gesonderten Kapitel die Rede sein wird. ASH (1949) wies nach, daß die Diagnosen verschiedener Psychiater aufgrund des gleichen Interviews nur ungenügend übereinstimmen. SIEGEL (1954) konnte zeigen, daß Kliniker und Nicht-Kliniker bei einem "Einfühlungstest" gleiche Ergebnisse erhalten. Kurze und längere Interviews haben keinen unterschiedlichen Einfluß auf Validität und Reliabilität der Ergebnisse (LEVIN 1954).

Eine zweite Eigenschaft, die ALLPORT anführt, ist die Ähnlichkeit des Beurteilers mit dem Beurteilten. Demnach könne man eine Eigenschaft, die man selbst in einem hohen Ausmaß hat, bei anderen besser beurteilen, weil Ähnlichkeit einen speziellen Fall der Erfahrung darstelle. Zieht man zu dieser Erklärung eine Annahme hinzu, welche projektiven Tests zugrunde liegt, so neigen wir dazu, unsere Besonderheiten in andere zu projizieren. Untersuchungen sprechen im allgemeinen für die Projektionshypothese (verg.u.a. BENDER und HASTORF 1950, CHANCE und MEADERS 1960, CROCKETT und MEIDINGER 1956, GORDON 1956/57, HALPERN 1955, LAING 1960, LUNDY 1956, NORMAN und AINSWORTH, SEARS 1936, SMITH 1960, SPANNER 1961). Gleichzeitig verdeutlichen diese Untersuchungen, daß die Tendenz, die eigene Eigenart in anderen Menschen wiederzufinden, nicht unabhängig von Eigentümlichkeiten der Situation, in welcher geurteilt wird, und von Besonderheiten des Beurteilers ist.

Besonders eindeutig wirkt die Tendenz zur Projektion bei der Beurteilung von Einstellungen und Motiven (vergl. u.a. BENDER und HASTORF 1950, PRECKER 1951). Hier wird deutlich, daß Beurteilungen mehr über den Beurteiler als über den Beurteilten aussagen.

Weiterhin soll sich der gute Menschenkenner nach ALLPORT dadurch auszeichnen, daß er Einsicht in seine eigenen Schwächen habe. Untersuchungen hierzu haben uneinheitliche Ergebnisse erbracht. ESTES (zitiert bei ALLPORT 1937, S.508) wies nach, daß Beurteiler nach einer Psychoanalyse nicht besser urteilten als solche, die keine psychoanalytische Behandlung hinter sich gebracht hatten. Andere Untersuchungen stellen heraus, daß mehr Projektionen vorkommen, wenn die Einsicht in die eigenen Schwächen fehlt, wodurch ALLPORT's Hypothese unterstützt wird. Die Einsicht in die eigenen Schwächen scheint dann vorteilhaft zu sein, wenn es sich um die Beurteilung von Einstellungen und Motiven handelt (DYMOND 1948, SEARS 1936, WEINGARTEN 1948/49).

Eine der wichtigsten Voraussetzungen für einen guten Menschenkenner ist nach ALLPORT die ästhetische Haltung die objektiv schwer zu erfassen ist. TAFT (1955) ließ Studenten verschiedener Fachrichtungen, darunter Musikstudenten, Urteile abgeben. TAFT fand keine Bestätigung von ALLPORT's Hypothese, seine Ergebnisse wiesen eher noch in die entgegengesetzte Richtung.

Weiterhin stellt ALLPORT die Hypothese auf, daß

Menschen, die differenzierter sind als man selbst, schwer zubeurteilen seien. Die kognitive Komplexität bzw. die Differenziertheit des Erkennens wurde von KELLY (1955), BIERI (1955) und LEVENTHAL (1957) untersucht. Die genannten Autoren bestätigen ALLPORT's Hypothese, daß die differenzierten Beurteiler zutreffender urteilen und weniger Neigung zur Projektion zeigen. Unter kognitiver Differenziertheit wird dabei die Anzahl unabhängiger Dimensionen verstanden, über die ein Beurteiler bei der Kategorisierung von Beobachtungsdaten aus der Personenwahrnehmung verfügt. Intelligente Menschen, die zur sozialen Distanz neigen, sind nach ALLPORT bessere Menschenkenner. Über den Zusammenhang zwischen Intelligenz und der Fähigkeit zur Menschenbeurteilung gibt es zahlreiche Untersuchungen (s. BRUNER und TAGUIRI, in LINDZEY 1954, S.644). Allgemein besteht eine geringe positive Beziehung zwischen der Validität der Urteile und der Intelligenz des Beurteilers. Möglicherweise wird die Beziehung unterschätzt, da die Untersuchungen in der Regel mit Studentenpopulationen durchgeführt wurden und von daher eine Varianzeinschränkung bezüglich des Merkmals Intelligenz besteht.

Übereinstimmung besteht zwischen ALLPORT und anderen Autoren in der Behauptung, daß kühle, introvertierte Personen bessere Urteile abgeben als sozial extravertierte Personen (ADAMS 1927, TAFT 1956, VERNON 1933). TAFT (1956) konnte zeigen, daß Menschen mit vielen Geschwistern relativ schlechte Beurteilungen abgeben. In diesen Zusammenhang gehört auch der Befund, daß Physiker und Meterologen

bessere Menschenkenner sind als Psychologen und Sozialwissenschaftler, denen man soziales Interesse und geringe Distanz zu anderen zuschreibt (LUNDY 1956).

2.6.2. Der Einfluß der beurteilten Eigenschaft

Es wäre noch zu erwähnen, daß man allgemein der Ansicht ist, daß die Richtigkeit der Beurteilung auch davon abhängt, welche Eigenschaften man in Betracht zieht. Man sagt, daß distale Eigenschaften zuverlässiger und valider beurteilt werden als proximale. Hierzu sei ein Zitat HOFSTÄTTER's (1957, S.98) angefügt: " Persönlichkeitseigenschaften, die sich im Stile des Verhaltens ausdrücken (Impulsivität und Gehemmtheit . . .) werden in der Regel richtiger beurteilt als Begabungs- und Wertausrichtungen." Wie bereits erwähnt, findet bei Werthaltungen und Einstellungen eine starke Projektion statt, so daß die Beurteilung über den Beurteiler mehr Auskunft erteilt als über den Beurteilten.

Die Kategorien, nach denen beurteilt wird, sind kulturell abhängig und nach Zweckmäßigkeit aufgebaut. Sie variieren in Kulturbereichen, Teilkulturen und in sozialen Schichten. Eine Untersuchung von ERTEL (1964) verdeutlicht, welche Beurteilungsgesichtspunkte in zwei verschiedenen Kulturen (bei Thai - ländern und Deutschen) verwendet werden. ERTEL benötigte "kulturunabhängige Reize" und wählte dazu eine Liste von Sinnbildern wie z.B. aufgehender Mond, schweres Gewicht, reife Frucht usw. Die Vpn

mußten diese Sinnbilder "zurückübersetzen" und angeben, was für ein Mensch zu dem jeweiligen Bild passe. Es sollten zwei konkrete Eigenschaften genannt werden, außerdem sollte angegeben werden, ob diese Eigenschaften wünschenswert oder nicht wünschenswert seien. Häufigkeitsvergleiche erbrachten, daß bei den Thailändern Reichtum und Armut eine viel wichtigere Bewertungskategorie darstellt als bei den Deutschen. Weiterhin wurden äußere Sicherheit und Unsicherheit, Glück und Mißgeschick, Vitalität und Nichtvitalität und gute äußere Erscheinung dreimal so häufig erwähnt wie bei den Deutschen, die zehnmal häufiger als die Thailänder Aufgeschlossenheit, Widerstandsfähigkeit, Anpassungsfähigkeit, Zielgerichtetheit, Entschlossenheit und Willensstärke nannten.

Aus dieser Untersuchung geht hervor, daß die Kategorien zumindest nach ihrer Häufigkeit in verschiedenen Kulturen verschieden sind. Allgemein kann man sagen, daß die Kategorien von Kultur zur Kultur, von Gruppe zu Gruppe und von Individuum zu Individuum verschieden sind. Jedes Individuum bevorzugt bestimmte Kategorien (etwa intelligent-nicht intelligent) zur Einstufung. Diese bleiben über Personen hinweg konstant und werden offenbar für lange Zeit beibehalten.

2.6.3. Der Einfluß der sozialen Beziehung zwischen Beurteiler und Beurteiltem

Weiterhin hängt "das Bild vom Anderen" mit der

speziellen sozialen Beziehung, die der Beurteiler zum Beurteilten hat, ab. Bei der Beurteilung von Müttern nach moralischen Qualitäten werden die meisten Kinder sehr gute Urteile abgeben, da die Mutter diejenige ist, die den Kindern das moralische Verhalten beibringt. Die sozialen Beziehungen und Rollen determinieren also sehr stark die Beurteilungen. Wer im Büro bescheiden und zaghaft ist, braucht das zu Hause noch lange nicht zu sein und umgekehrt. Wenn es in einer Schulklasse Ärger gegeben hat und ein Schüler von dem Lehrer nach dem Grund gefragt wird, so wird das eventuelle Schweigen des Schülers vollkommen verschieden ausgelegt: Vom Lehrer als verstockt, frech, aufsässig, von den Mitschülern als tapfer und kameradschaftlich. Die gleiche Verhaltenssequenz kann zu völlig verschiedenen Beurteilungen führen. Diese Befunde sprechen gegen die Verwendung allgemeiner Beurteilungen.

Fassen wir die Menschenbeurteilung als eine Grundlage sozialen Verhaltens auf, so tritt die Frage nach der Richtigkeit der Urteile in den Hintergrund. Man könnte sagen, daß die Beurteilung des Mitmenschen als soziales Verhalten im sozialen Zusammenleben gelernt wird. Ein Urteil wird nur dann bestätigt (verstärkt), wenn man sich an die in der Gruppe herrschenden Regeln hält. Das Urteil: Mein Vater ist dumm, kommt womöglich gar nicht erst auf, da es bestraft wird. Da man Beurteilungen in der Regel nur über Abwesende abgibt, tritt keine direkte Rückmeldung auf, sondern nur die der Gruppe. Hierdurch erklären sich möglicherweise stereotype Urteile. In diesem Zusammenhang sei noch ein interessanter Versuch MITTENECKERs (1960) angefügt. Zwei Versuchs-

personengruppen hatten nach Photos die Intelligenz von Personen einzustufen. In der Experimentalgruppe erhielten die Vpn nach jeder Einstufung eine Information über die Richtigkeit ihres Urteils, die Kontrollgruppe erhielt keine Rückmeldung. Die Richtigkeit der Urteile in der Experimentalgruppe nahm zu, schon nach einer kurzen Übungsperiode erhöhte sich die Trefferzahl. Man könnte dieses Ergebnis dahin-gehend interpretieren, daß man auf Rückmeldung angewiesen ist; man wird gerade zu daran gehindert, richtige Urteile abzugeben, wenn keine Rückmeldung erfolgt, stattdessen entstehen Vorurteile. Der Versuch MITTENECKER's hätte noch um eine weitere Experimentalgruppe ergänzt werden müssen, um zu überprüfen, ob falsche Rückmeldung zu konsistent falschen Urteilen führt.

2.7. Reliabilität und Validität von Lehrerurteilen, sowie sie in Schulnoten zum Ausdruck kommen

Das Kriterium für die Leistungsbeurteilung in der Schule sind die Schulnoten. Bei den Schulnoten handelt es sich um ein Schätzverfahren nach der Methode des subjektiven Eindrucks, also um eine individuelle Beurteilung, die sehr unsicher ist. REMMERS (in GAGE 1970, S.954) hat die folgenden Fehlerfaktoren zusammengestellt, die bei Schätzurteilen auftreten können:

a) Zeitwahlfehler: Unterrepräsentation von relevanten Verhaltensweisen, die möglicherweise zu wenig vorkommen, um in einer Zeitstichprobe enthalten zu sein.

b) Erfahrungsfehler: Häufig sind Junglehrer weniger sicher in ihrem Urteil; so können sie z.B. eher dazu neigen, mittlere Urteile abzugeben im Vergleich zu erfahrenen Lehrern.

c) Merkmalsverzerrung: Durch starke Gewichtung einzelner, hoch korrelierender Verhaltensweisen werden diese insgesamt zu stark betont.

d) Schätzurteilsverzerrung: Hierunter fallen Reaktionstendenzen, die individuell bedingt sind, wie der schon erörterte Halo-Effekt, Wahrnehmungsselektionen durch Vorurteile, die Tendenzen zur mittleren bzw. extremen Beurteilung.

Die langfristige Leistungsvorhersage, wie sie die Auslese von Schülern in unserem dreigliedrigen Schulsystem mit punktueller Selektion verlangt, setzt voraus, daß die Schulleistungen der Kinder über die Zeit hinweg einigermaßen konstant bleiben. Die Noten der Schüler dürften sich im Laufe der Zeit nicht allzu stark verändern. Diesem Aspekt entspricht das Konzept der Reliabilität. Wären Schulnoten also reliabel, so müßten sich zukünftige Noten aus früheren vorhersagen lassen. RANK (1962) hat für die Fächer Deutsch und Mathematik die Noten vom 1. bis zum 8. Schuljahr zueinander in Beziehung gesetzt und festgestellt, daß nur in der Grundschule von einigermaßen brauchbaren Reliabilitätskoeffizienten die Rede sein kann. Die Koeffizienten liegen um .60 bis .70; von Jahr zu Jahr liegen die Werte um .80, innerhalb eines Schuljahres um .90. Bei diesen hohen Koeffizienten ist allerdings zu bedenken, daß sie z.T. auf den sog. Halo-Effekt zurückgehen. In der Regel ist es ja derselbe Lehrer,

der die Noten in den verschiedenen Fächern und für die verschiedenen Zeugnisse erteilt, und es besteht eine Tendenz, einmal gefaßte Urteile nur noch wenig zu korrigieren. TENT (1969) fand, daß Zeugnisnoten innerhalb der Schuljahre systematisch reliabler sind als für gleichlange Abschnitte von einem Schuljahr zum nächsten. Das hängt u.a. damit zusammen, daß die Lehrer häufiger mit dem Schuljahr als innerhalb des Schuljahres wechseln.

Für Gymnasiasten erhielt RANK (1962) beim Vergleich der Leistungen vom 4. zum 8. Schuljahr Koeffizienten von nur .2 und zwischen dem 1. und 8. Schuljahr von .1. In diesen Werten deutet sich bereits an, wie gering die Möglichkeit ist, spätere Leistungen auf dem Gymnasium aufgrund der Grundschulleistungen vorherzusagen.

WEINGARDT (1964) untersuchte die Stabilität der Schulnoten erfolgreicher Schüler in den einzelnen Fächern über die gesamte Gymnasialzeit, also vom 5. bis zum 13. Schuljahr. Die Note mit der größten Stabilität ist die Lateinnote mit einem Koeffizienten von .51. Eine ähnliche Größenordnung erreichen nur noch Leibeserziehung und Englisch. Alle anderen Koeffizienten liegen um .30. Es ist auch nach diesen hier nur angedeuteten Ergebnissen kaum möglich, eine Vorhersage mit dem von unserem Schulsystem her geforderten Maß an Reliabilität zu machen. TENT (1969) konnte in diesem Zusammenhang die Auswirkung des Lehrerwechsels genauer kontrollieren. Für die Fächer Deutsch, Mathematik und Fremdsprache wurde die Übereinstimmung der

Noten bei bzw. ohne Lehrerwechsel vom 8. zum 9. Schuljahr festgestellt. Die Stabilität der Schulnoten ist ohne Lehrerwechsel größer, mit einem Lehrerwechsel wird sie stark beeinträchtigt.

Wenn über die Validität von Schulnoten gesprochen wird, so ist die sog. Vorhersagevalidität relevant, d.h. inwieweit Schulnoten zur Vorhersage der Eignung für weiterführende Schulen herangezogen werden können. Bei der Betrachtung von Schulnoten fällt dieser Aspekt teilweise mit dem der Reliabilität zusammen. Die Höhe der zu fordernden Gültigkeit ist abhängig vom Zweck. Bei individueller Begutachtung und bei Auslese ist eine relativ hohe Gültigkeit erforderlich, die Angaben schwanken zwischen .60 bis .90. Für Einzelprobleme kann schon eine Gültigkeit von .30 genügen.

Daß in der Literatur sehr niedrige Validitätskoeffizienten angegeben werden, liegt u.a. an den folgenden Bedingungen:

Der Validitätskoeffizient ist als Korrelationskoeffizient populationsabhängig. So tritt z.B. bei homogenen Populationen ein niedrigerer Koeffizient auf als zu erwarten ist. Eine solche Homogenisierung kommt durch Selektion von Schülern in weiterführenden Schulen zustande. Die Höhe des Koeffizienten ist auch abhängig vom Kriterium. Durch die"Heterogenisierung" des Kriteriums kann mehr Varianz des vorherzusagenden Merkmals erfaßt werden. Das Kriterium müßte dann mit einer möglichst heterogenen Testbatterie vorhergesagt werden. Als Maßzahl für die

Validität der Testbatterie dient dann nicht mehr der einfache, sondern der multiple bzw. kanonische Korrelationskoeffizient.

Im folgenden soll dargestellt werden, welchen Voraussagewert die benoteten Prädiktoren trotz der möglicherweise vorhandenen Fehlerfaktoren bei der Beurteilung besitzen. Um die Gültigkeit der konventionellen Ausleseverfahren zu überprüfen, führte SCHULTZE (1964) an 5114 Schülern eine Untersuchung durch. Er berechnete Rangkorrelationen mit dem Schulerfolg. Die folgende Tabelle (entnommen aus SCHULTZE 1964, Tab. 21, S.50) zeigt die Validitätskoeffizienten der benutzten Prädiktoren nach vier und fünf Jahren.

Prädiktoren	nach 4 Jahren	nach 5 Jahren
Grundschulgutachten	. 28	. 25
Grundschulzeugnis:		
Rechnen	. 28	. 28
Deutsch	. 29	. 31
Schriftliche Aufnahmeprüfung (gesamt)	. 35	. 33
Diktat	. 28	. 26
Aufsatz	. 25	. 26
Rechnen	. 23	. 23

Keiner der Prädiktoren hat ausreichendes Gewicht, um eine gesicherte Vorhersage zu ermöglichen. Die konventionelle Aufnahmeprüfung liefert keine zusätzliche Information zu den Grundschulgutachten. Weiterhin untersuchte SCHULTZE die Gültigkeit der Gesamtnote

der schriftlichen sowie der mündlichen Aufnahmeprüfung in Deutsch und Rechnen an einer hessischen Schülerstichprobe (N=430). Die Ergebnisse sind in der folgenden Tabelle (entnommen SCHULTZE 1964, Tab.23, S. 53) dargestellt:

r_{stc}	nach Jahren 1	2	3	4	5	6
Schriftl.Aufnahmeprüfung	.16	.31	.35	.39	.36	.33
Schriftliche und mündl. Aufnahmeprüfung	.13	.22	.30	.31	.29	.28

Es zeigt sich in dieser Untersuchung, daß aus der Hinzunahme der Beurteilung der mündlichen Leistung, entgegen der häufig von Pädagogen vertretenen Meinung, eine Verringerung der Validität resultiert.

Eine Arbeit, in der Tests als Prädiktoren zu der Aufnahmeprüfung hinzugezogen wurden, lieferte HITPASS (1961). Die Gültigkeit der beiden Verfahren bei der Vorhersage späteren Schulerfolgs wurde festgestellt, indem man die Korrelationen über 11 Jahre hinweg verfolgte. In die Validitätsprüfung wurden einbezogen:

a) schriftliche Aufnahmeprüfung: Diktat, Aufsatz und Rechnen
b) eine Begabtentestreihe, u.a. der H W Y-Test
c) der Arbeitsversuch von PAULI.

In der folgenden Tabelle sind die biserialen Validitätskoeffizienten dargestellt (entnommen HITPASS 1961, Tab.6, S.377):

Prädiktoren	r_{bis}
Aufnahmeprüfung (gesamt)	-0.09
Diktat	-0.25
Aufsatz	0.09
Rechnen	0.00
Testverfahren (gesamt)	0.64
HWY-Test	0.78
PAULI-Test	0.38

Der Validitätsvergleich von Aufnahme- und Testprüfung fällt eindeutig zugunsten der Testprüfung aus. Die Aufnahmeprüfung besitzt keinerlei Gültigkeit. Nur 14 % der Sextaner gelangten in direktem Aufstieg zum Abitur.

Auch Untersuchungen von GEBAUER (1964) belegen, daß Tests im Vergleich zur konventionellen Aufnahmeprüfung einen weit höheren prognostischen Wert haben.

Um von den wenig validen Aufnahmeprüfungen wegzukommen, überprüfte TENT (1969) das Leistungsprüfsystem (LPS) von HORN als Ausleseverfahren, indem er den Gesamtpunktwert dieses Tests zu den ein Jahr nach dem Übergang zur weiterführenden Schule erzielten Schulnoten in Beziehung brachte (N=942). Aufgrund einer kanonischen Korrelationsanalyse kann man mit einer Vorhersagevalidität von etwa r= 0.7 rechnen. Berücksichtigt man die Bedeutung der Vorhersage, so ist auch eine Validität dieser Größenordnung noch als zu gering anzusehen. Allerdings ist die Vorhersagegenauigkeit des Tests immer noch wesentlich größer als die von den konventionellen Verfahren. TENT (1969) untersuchte daraufhin die Zulänglichkeit der Vorhersage durch Tests. Man kann die

Zulänglichkeit eines Verfahrens verstehen als das Verhältnis der aufgeklärten zu der überhaupt aufklärbaren Varianz. Letztere ist bestimmt durch die Reliabilitäten von Kriterium und Prädiktor. Einem solchen Maß entspricht der doppeltminderungskorrigierte Validitätskoeffizient. Berechnet man diesen Koeffizienten für die Vorhersage von Schulnoten aufgrund des LPS-Gesamtpunktwertes, so ergibt sich ein Wert von r= o.99. Damit ist im Rahmen des Systems die maximal mögliche Vorhersage erreicht. Weiter an der Verbesserung von Vorhersageinstrumenten zu arbeiten, ist daher wenig sinnvoll, d.h. zunächst müßte das Beurteilungssystem geändert werden.

Die genannten Untersuchungen beschäftigen sich mit der Auslese auf weiterführende Schulen, während die folgende Arbeit von ORLIK (1967) sich mit der Auslese aufgrund des Reifezeugnisses befaßt, und zwar im Hinblick auf die nachfolgende Studienförderung, wobei das Reifezeugnis als Prädiktor dient. Um ein differenzierteres Leistungsmaß als den Reifezeugnisgesamtdurchschnitt zu bekommen, entwickelte ORLIK ein Reifezeugnisprofil, das Begabungsrichtungen und Leistungsschwerpunkte deutlich machen soll. Hierbei bilden acht ausgewählte Fächer die Abszisse und die standardisierten Zensurenwerte die Ordinate des Profils. Diese Profile vermögen nach ORLIK zwischen den verschiedenen Studienrichtungen besser zu diskriminieren als z.B. der Intelligenz-Struktur-Test (IST) von AMTHAUER. Validitätsmaße, die eine Einordnung der Ergebnisse erlauben, legte ORLIK jedoch nicht vor.

In der folgenden Erörterung soll noch einmal auf

die Frage der Auslese zu weiterführenden Schulen eingegangen werden, hier jedoch unter dem Aspekt, daß alle Schüler nach ihrem Übertritt überprüft werden. Von INGENKAMP (1963) liegt hierzu eine Untersuchung an über 1.000 Schülern der 6.Grundschulklasse in Berlin vor. Die Testverfahren 'Frankfurter Analogietest, Rechtschreib-, Lese-und Rechentest' legte er den Schülern vor. Der Übertritt selbst, entweder in den praktischen, technischen oder wissenschaftlichen Zweig, richtete sich jedoch nach dem Grundschulgutachten oder der Wahl der Eltern. Es zeigte sich, daß die Schüler der weiterführenden Schulzweige einen signifikant höheren Mittelwert bezüglich der verwendeten Tests aufwiesen. Streuungsunterschiede konnten jedoch nicht gesichert werden. Die Leistungsverteilungen der einzelnen Schultypen überschnitten sich erheblich, so daß auf die Weise eine eindeutige Trennung zwischen den Schultypen nicht möglich ist. Die Trennung, die durch die Gutachten und die Elternwahl erfolgte, bezieht sich nicht auf eine begabungsgerechte Auslese, sondern auf den augenblicklichen Leistungsstand der Schüler und die subjektive Beurteilung der Lehrer, die häufig nur an einem klasseninternen Maßstab orientiert ist.

In den angeführten Untersuchungen zeigte sich, daß ein Schulsystem mit konventioneller Leistungsbeurteilung und punktueller Selektion als integrierter Bestandteil des Gesamtsystems nicht befriedigend funktionieren kann. Hierzu TENT (1969, S. 55):" Die Eigenart des urteilenden Menschen setzt der üblichen Notengebung in den Schulen dieselben Grenzen wie anderen Schätzurteilen dieser Art - die Urteile hängen ab von einer notwendigerweise selektiven Wahrnehmung,

der Ungenauigkeit seiner Beobachtung, von Behalten und Vergessen und von seiner möglicherweise mangelhaften Sensitivität für das pädagogisch und psychologisch Relevante" (vgl. TENT & REMMERS 1969, Abschnitt 1).

2.8. Eröterungen zum Problem der kognitiven Komplexität

In neueren Untersuchungen zur interpersonellen Wahrnehmung tritt die Differenziertheit der Personwahrnehmung mehr und mehr in den Blickpunkt des Forschungsinteresses (vgl.u.a. ADINOLFI 1971, BIERI et al.1966; KROHNE 1973). Urteilsprozesse sollen deshalb durchschaubarer gemacht werden, um systematisch variierende Beurteilerfehler zu minimieren, insbesondere beim klinisch arbeitenden Psychologen. KROHNE (1973) weist auf die Notwendigkeit der Trennung zwischen Urteilsgenauigkeit und Urteilsvalidität hin:

" Als zentrales Problem der Urteilsbildung wurde in diesem Zusammenhang das der Genauigkeit (accuracy) des Urteils erkannt (vgl.u.a. CLINE 1964, FANCHER 1966, 1967, WARR und KNAPPER 1968). Empirischer Indikator von Urteilsgenauigkeit ist das Ausmaß, in dem ein Beurteiler Verhaltensdaten einer zu beurteilenden Person auf der Grundlage bestimmter Informationen treffsicher vorhersagen kann, wobei Vorhersage nicht nur zeitlich, sondern auch querschnitthaft gemeint ist. Von diesem Aspekt zu trennen ist der der Urteilsvalidität, d.h. das Ausmaß, in dem ein Beurteiler Verhaltensdaten einer

Person in konsistenter und kommunizierbarer Weise zu einer Beschreibung bzw. Theorie dieser Person integrieren kann (FANCHER 1967). Nach Überlegungen TAFTS (1955) können diese beiden Aspekte der Personalbeurteilung, entgegen naheliegenden Vermutungen, unkorreliert sein."

Ausgehend von KELLY's (1955) und BIERI's et al. (1966) Modell der kognitiven Strukturiertheit erarbeiteten HARVEY, HUNT und SCHRODER (1961) und SCHRODER, DRIVER und STREUFERT (1967) eine allgemeine Theorie zur menschlichen Informationsverarbeitung. In diesem Modell werden zwei Komponenten am Informationsverarbeitungsprozeß unterschieden, wie KROHNE (1972) in einer Antrittsvorlesung komprimiert und zusammenfassend darlegte: " Die Informationsaufnahme und die nachfolgende Organisation dieser aufgenommenen Information. Die Aufnahmekomponente wird hierbei als eine Vorrichtung gedacht, die eingehende Reize nach bestimmten Attributen oder Dimensionen ordnet, d.h. sie kategorisiert, und ihnen dann auf diesen Dimensionen einen bestimmten Wert zuweist,d.h. sie skaliert. Diese beiden Bereiche der Informationsaufnahme werden auch als Differenzierung und Diskriminierung bezeichnet. Die Organisationskomponente ist in ihrer komplexesten Ausprägung als ein System von Regeln oder als ein Programm konzipiert, das auf der Grundlage ein und desselben objektiven Reizes und ein und desselben Satzes von Skalenwerten dieses Reizes auf verschiedenen Dimensionen unterschiedliche Sichtweisen (Konzepte) über diesen Reiz erzeugen kann. Das Programm gewichtet dabei die einzelnen Skalenwerte auf den verschiedenen zur Katego-

risierung herangezogenen Dimensionen in jeweils unterschiedlicher Weise und integriert diese Werte dann je nach ihrem Gewicht zu verschiedenen "Perspektiven" bzw. Konzepten dieses Reizes im Denken, Urteilen oder Handeln. Dieser Teil der Informationsverarbeitung wird als Integration bezeichnet. Individuen unterscheiden sich in der Art, wie sie Informationen aufnehmen und organisieren. Analog zu den dargestellten drei Komponenten der Informationsverarbeitung unterscheiden sich Menschen erstens nach der Anzahl der Dimensionen oder Skalen, auf denen sie einen objektiven Reiz zu erfassen vermögen (Differenzierung), zweitens nach der Anzahl der Skalenwerte, die sie einem Reiz auf jeweils einer dieser Dimensionen zuweisen können (Diskriminierung), und drittens nach der Anzahl und der Strukturiertheit der kombinierenden Regeln, nach denen sie diese einzelnen Skalenwerte zu einem oder mehreren Konzepten eines Reizes verbinden (Integration). Der Grad der Leistungsfähigkeit dieser drei Komponenten der Informationsverarbeitung bestimmt das Niveau der Informationsverarbeitungsfähigkeit eines Individuums. SCHRODER et al. nennen diese Vaiable "kognitive Komplexität" und bezeichnen ihre Pole als konkretes bzw. simples Niveau (für einfache Formen der Informationsverarbeitung) und als abstraktes bzw. komplexes Niveau (für hochstruktuierte Informationsverarbeitung).

Empirische Untersuchungen (vgl.KROHNE in Vorb.) haben gezeigt, daß ein Anstieg in kognitiver Komplexität mit Veränderungen in folgenden Verhaltens-

weisen einhergeht:

1) Anstieg in subjektiver Ungewißheit und Konfliktoleranz bei der Verarbeitung komplexer Information.

2) Verlängerung der Entscheidungszeit von der Konfrontation mit einem Problem bis zur endgültigen Lösung.

3) Erhöhung des Ausmaßes der Informationssuche vor einer Entscheidung.

4) Zunahme kreativer Problemlösungen.

5) Verstärkung des Bewußtseins der Selbsterzeugung von Wissen und der Selbstverursachung von Handlungen, d.h., eine Veränderung auf der Dimension Heteronomie - Autonomie.

6) Abbau von ethnozentrischen Haltungen sowie autoritären und dogmatischen Stilen.

7) Aufgabe einer streng hierarchisch organisierten, wenig flexiblen Gruppenstruktur zugunsten einer funktionalen Rollenverteilung, einschließlich der Führerrolle, d.h. einer Organisation der Gruppe nach jeweiligen Entscheidungszielen."

2.9. Folgerungen und Hinweise zum Untersuchungsablauf

Überblickt man die zahlreichen Untersuchungen zur Beurteilung der Mitmenschen, so tritt in allen Untersuchungen das Problem der Validierung der Beurteilungen auf, das sehr unterschiedlich und meist unbefriedigend gelöst wurde. Soweit man die verschiedenen Arbeiten überhaupt vergleichen kann, ergibt sich folgendes Bild: Die Fähigkeit, andere zu beurteilen, variiert zwar interindividuell, ist aber spezifisch in bezug auf den Pb und das Merkmal.

Deshalb möchte diese Arbeit dem einzelnen Lehrer, der aufgrund seiner beruflichen Situation ständig gezwungen ist, über Schüler weitreichende Beurteilungen abzugeben, verschiedenartige Zusammenhänge bei seiner Urteilsbildung durchschaubarer machen und ihm insbesondere aufzeigen, welche Bewertungsaspekte in seine Notengebung einfließen.

Im ersten empirischen Teil dieser Arbeit soll untersucht werden, ob sich mit Hilfe der Methode der Clusteranalyse Beurteilertendenzen bei Lehrern bezüglich der Leistungs- und Persönlichkeitseinschätzung ihrer Schüler aufdecken lassen. Es soll herausgefunden werden, ob, aufgrund welcher und wievieler Urteilsdimensionen zur aktuellen Schülerbeurteilung Lehrer in ihren Klassen unterscheidbare und konsistente Schülergruppen bilden. Dabei wird das HOFER'sche Beurteilungsschema zur aktuellen Schülerbeurteilung herangezogen, das neben Leistungs- auch Persönlichkeitseinschätzungen vom Lehrer verlangt. Die Anzahl unabhängiger Dimensionen, die signifikant zur Gruppentrennung beurteilter Schüler in einer Klasse beiträgt, könnte sich u.U. als Indikator für kognitive Differenziertheit bei einem Lehrer herausstellen.

Um eine kritische Einstellung einzelner Lehrer angesichts ihrer Beurteilungsprozesse zu fördern, wenden wir uns der Frage zu, ob und inwieweit sich die undifferenzierte Globalkritik von Lehrern gegenüber Schülern, eine Globalkritik, wie sie sich in den Schulnoten manifestiert, auf das implizite Persönlichkeitsmodell der Lehrer zur aktuellen Schülerbeurteilung (nach HOFER) reduzieren läßt. Es soll überprüft wer-

den, inwieweit und mit welcher Gewichtung bei einzelnen Lehrern die Urteilsdimensionen Diszipliniertheit vs. Undisziplin[illegible]rtheit, geistige Regsamkeit, bescheidene Zurückhaltung vs. Geltungsstreben, soziale Aufgeschlossenheit vs. Verschlossenheit und Empfindsamkeit vs. Robustheit in die Zeugnisnoten unterschiedlicher Fächer einfließen. Die Bearbeitung dieser Frage erscheint uns insofern von Belang, da sie ggf. Aufschlüsse über die Gewichtung unterschiedlicher Bewertungskriterien im Urteilsprozeß bei einzelnen Lehrern gestattet.

II EMPIRISCHER TEIL

3. Fragestellungen

1. Lassen sich bei einzelnen Lehrern Beurteilertendenzen nachweisen, derart, daß unter Zugrundelegung der fünf HOFERschen Beurteilungsdimensionen Arbeitsverhalten, soziale Aufgeschlossenheit vs. Verschlossenheit, gei ti- ge Regsamkeit, bescheidene Zurückhaltung vs. Geltungs- streben und Empfindsamkeit vs. Robustheit in Klassen konsistente und unterscheidbare Schülergruppen auffind- bar sind?

2. Durch welche und wieviele Urteilsdimensionen lassen sich etwaige Schülergruppierungen in den einzelnen Klassen unterscheiden?

3a. Lassen sich bei einzelnen Lehrern Zusammenhänge zwischen den fünf HOFERschen Beurteilungsdimensionen einerseits und den Zeugnisnoten in bestimmten Fächern andererseits nachweisen?

3b. Ändern sich die in 3a angesprochenen Zusammenhänge, je nachdem ob Jungen oder Mädchen beurteilt werden?

Falls sich die in Frage 3 angesprochenen Zusammenhänge auf- zeigen lassen, erscheint es sinnvoll, die folgenden weiter- führenden Fragen zu bearbeiten:

4a. In welcher Rangfolge und mit welchem Gewicht sind die fünf Urteilsdimensionen am Gesamtzusammenhang mit den Schulnoten beim einzelnen Lehrer beteiligt?

4b. Lassen sich beim einzelnen Lehrer entsprechend Frage 4a Unterschiede nachweisen, je nachdem, ob er Jungen oder Mädchen beurteilt?

4. Durchführung der Untersuchung

4.1 Untersuchungsmaterial

Im einleitenden Kapitel dieser Arbeit wurde bereits dargelegt, daß Lehrer feste Vorstellungen über Eigenschaftszusammenhänge bei Schülern besitzen. Diese Eigenschaftszusammenhänge beschreibt HOFER (1969) durch die Faktoren Diszipliniertheit vs. Undiszipliniertheit, Zurückhaltung vs. Geltungsstreben, geistige Regsamkeit, soziale Aufgeschlossenheit vs. Verschlossenheit und Empfindsamkeit vs. Robustheit. HOFER (1969, S. 30-32) fand aus einer Primärliste von 191 Eigenschaftswörtern diejenigen Eigenschaften heraus, die Lehrer häufig zur Beurteilung von Schülern verwenden und die das Urteilsverhalten von Lehrern umfassend (repräsentativ) beschreiben. Außerdem schloß er homonyme und synonyme Eigenschaftsbegriffe aus. Sein Beurteilungsbogen zur aktuellen Beurteilung der Schüler umfaßt 25 Eigenschaften, die in gleichem Ausmaß positive wie negative Aspekte einschließen. Zwischen den polar angeordneten Eigenschaften dieses Bogens (Polaritätenprofil) bestehen sieben Möglichkeiten der Abstufung. Der Eigenschaftsbogen mit vorgegebener Instruktion ist dem Anhang, (S.123) beigefügt.

Neben den Beurteilungen anhand des Eigenschaftsbogens faßten wir diejenigen Noten ins Auge, die von den an der Untersuchung beteiligten Lehrern im letzten Zeugnis erteilt wurden.

4.2. Datenerhebung

An der Datenerhebung, die in der Zeit von Januar bis März 1972 stattfand, nahmen Schulen aus Dortmund, Bochum, Düsseldorf und Osnabrück teil. Aus Gründen der Anonymität wurden die beteiligten Schulen und Klassen nicht näher benannt.

In Tab. 1 sind die an der Untersuchung beteiligten Jahrgänge und Schultypen aufgeführt.

Tab. 1
Zusammensetzung der Stichprobe

Schulart	Klasse (Jahrgang)				
	7	8	9	10	Insgesamt
Hauptschule	6	2	3	1	12
Realschule	3	4	2	-	9
Sonderschule für Lernbehinderte	3	8	4	-	15
Insgesamt	12	14	9	1	36

Das Alter der beteiligten Lehrer und Lehrerinnen schwankte zwischen 22 und 68 Jahren.
Tab. 2 ist die Altersverteilung zu entnehmen.

Tab. 2
Altersverteilung der an der Untersuchung beteiligten Lehrer

Alter	20-30	31-40	41-50	über 50	Gesamt
Lehrer	3	9	7	2	21
Lehrerinnen	6	6	3	-	15

Die Lehrer und Lehrerinnen wurden gebeten, jeden Schüler bzw. jede Schülerin ihrer jeweiligen Klasse anhand der vorgegebenen Polaritäten einzustufen.

Aus Tab. 3 geht hervor, in welchen Fächern die letzten Zeugnisnoten bei welchen Lehrern bzw. welchen Klassen erhoben werden konnten. Die infragekommenden Fächer sind mit einem Stern gekennzeichnet.

Tabelle 3

Überblick über die Fächer, in denen Noten erhoben wurden

Lehre.	schl.	Schule	Jahrg.	N	Deutsch	Mathematik	Biologie	Physik	Englisch	Geschichte	Erdkunde
1	m	S	7	12	*	*					
2	w	S	8	14	*	*					
3	m	S	8	18	*	*					
4	m	S	7	14							
5	w	S	9	17	*	*					
6	w	S	8	10	*	*					
7	w	S	9	15	*	*					
8	m	S	8	15	*	*					
9	m	S	8	14							
10	w	S	8	15							
11	w	S	7	24	*	*					
12	m	S	8	13	*	*					
13	w	S	9	16	*	*					
14	m	S	8	13	*	*					
15	m	S	9	18	*	*					
16	m	H	10	26			*		*		
17	m	H	7	26			*		*		
18	w	H	9	24	*						
19	w	H	8	29	*					*	
20	w	H	9	22	*					*	
21	w	H	7	36	*		*				
22	m	H	7	36	*						*
23	w	H	7	25	*						
24	w	H	7	21	*						
25	m	H	7	23	*		*				
26	m	H	9	22		*					
27	m	H	8	22	*	*					
28	w	R	7	37	*				*		
29	m	R	8	25		*		*			
30	m	R	7	37		*		*			
31	m	R	9	31					*	*	
32	m	R	9	21		*		*			
33	m	R	7	35					*		
34	m	R	8	27							
35	m	R	8	30							
36	w	R	8	28			*		*		

Legende:

N = Anzahl der Schüler pro Klasse
S = Sonderschule für Lernbehinderte
H = Hauptschule
R = Realschule

4.3. Datenanalyse

4.3.1. Faktorenanalyse (R-Analyse) des Polaritätenprofils

Bevor die erste Untersuchungsfrage durch Clusteranalysen bearbeitet wird, sollen die HOFERschen Polaritäten, anhand derer die Lehrer jeden ihrer Schüler zu beurteilen hatten, faktoren- und itemanalysiert werden. Der Zweck der Faktorenanalyse an dieser Stelle ist es, die 25 polaren Eigenschaften (Items) zu "Subtests" zusammenzufassen. In die Faktorenanalyse mit anschließender Varimax-Rotation gehen die 25 Polaritäten des Schülerbeurteilungsbogens als Variablen und 810 beurteilte Schüler (aus 36 Klassen) als Meßwertträger ein. Tab. 4a gibt das von uns zu interpretierende Ergebnis der R-Analyse aus dem Rechenprotokoll wieder. Der Ausdruck zeigt, daß 5 Eigenwerte (λ_j) über 1 liegen. Außerdem ist dem Ausdruck zu entnehmen, daß die Kommunalitäten auffallend hoch ausfallen (zwischen .51 und .76). Zur Bestimmung der Anzahl der zu interpretierenden Faktoren ziehen wir zwei häufig angewandte Kriterien heran (vergl. ÜBERLA, 1968, S. 127):

1. Es werden die Faktoren interpretiert, die ein $\lambda_j \geq 1$ aufweisen;
2. Als zweites Kriterium wird die graphische Darstellung des Eigenwertabfalles herangezogen:

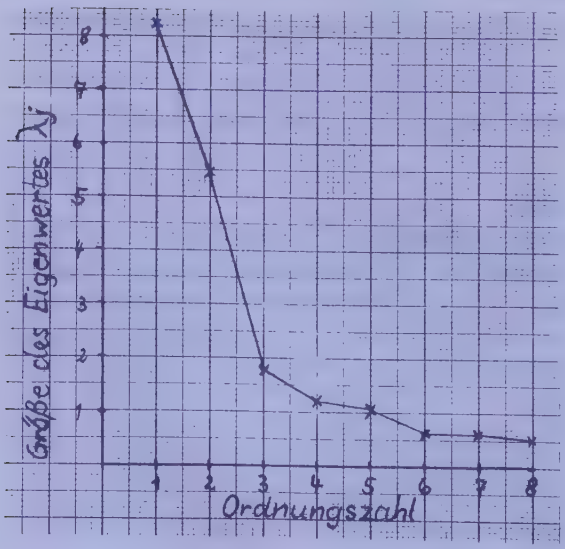


Abb. 1
Eigenwertverlauf zur Faktorenanalyse der Polaritäten

Tab. 4a

Ausdruck des Rechenprotokolls der Faktorenstruktur der Polaritäten

ROTATION MIT DELTA = 0,010 BETA = 0,001

Polaritäten (Items)

Faktoren	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-0.56	-0.19	0.17	0.05	-0.23	0.22	-0.03	-0.25	0.83	-0.20	0.21	0.83	-0.71
2	-0.27	-0.72	-0.44	0.49	-0.01	0.33	0.68	0.68	-0.06	-0.18	0.07	0.17	0.24
3	-0.51	-0.15	-0.00	0.42	-0.89	0.71	0.21	0.04	0.14	0.05	0.87	0.21	-0.33
4	-0.20	-0.13	0.26	-0.01	-0.08	0.05	0.33	-0.38	0.22	-0.24	0.08	0.09	-0.05
5	0.03	0.28	0.65	-0.47	0.04	-0.04	-0.22	0.07	0.03	0.74	0.05	0.14	0.24

Faktoren	14	15	16	17	18	19	20	21	22	23	24	25
1	0.31	-0.60	0.24	0.40	-0.39	0.83	0.75	0.34	0.33	0.72	0.21	-0.04
2	-0.19	0.23	-0.71	-0.33	0.73	-0.12	-0.18	0.43	-0.28	-0.06	0.21	0.75
3	0.06	-0.13	-0.09	-0.03	0.03	0.11	0.32	0.62	0.09	0.16	0.24	0.20
4	0.72	-0.27	0.41	0.65	-0.15	0.24	0.15	0.09	0.53	0.41	0.70	-0.15
5	0.01	0.08	0.12	-0.11	0.15	-0.02	-0.18	-0.07	-0.44	-0.10	-0.00	-0.18

EPS = 0.010 LAMBDA = 1.000

EIGENWERTE

1	2	3	4	5	6	7	8	9	10	11	12	13
8.17	5.48	1.63	1.16	1.01	0.66	0.63	0.57	0.54	0.49	0.43	0.41	0.38

14	15	16	17	18	19	20	21	22	23	24	25
0.37	0.35	0.35	0.33	0.30	0.29	0.27	0.24	0.21	0.20	0.18	0.16

KOMMUNALITAETEN

1	2	3	4	5	6	7	8	9	10	11	12	13
0.68	0.68	0.71	0.63	0.85	0.67	0.66	0.67	0.76	0.69	0.82	0.79	0.73

14	15	16	17	18	19	20	21	22	23	24	25
0.65	0.51	0.76	0.71	0.74	0.77	0.75	0.71	0.68	0.72	0.64	0.66

Auf der Abzisse des Koordinatenkreuzes werden die Ordnungszahlen der Eigenwerte abgetragen, auf der Ordinate die Größe der Eigenwerte, wie in Abb. 1. Die Anzahl zu interpretierender Faktoren ergibt sich, indem man eine Gerade zeichnet, die bei dem Eigenwert mit der höchsten Ordnungszahl beginnt. Alle Eigenwerte, die auf dieser Geraden liegen, gehören zu Faktoren, die nicht interpretiert werden, da sie als Zufallsfaktoren bezeichnet werden können.

Diesen beiden Kriterien zufolge bietet sich, wie schon bei HOFER (1969), eine Fünfer-Struktur als beste Lösung an. Als Ergebnis der R-Analyse können die 25 polaren Eigenschaften zu den in Tabelle 4b aufgeführten 5 Faktoren zusammengefaßt werden. Tabelle 4b enthält außerdem die Trennschärfekoeffizienten r_{it} der einzelnen polaren Eigenschaften (Items) bezogen auf den jeweiligen als Subtest betrachteten Faktor, zudem die Mittelwerte $\bar{X}$ der Polaritäten. Bis auf Faktor IV, den wir als "Normenkonformität" interpretierten, konnte die Benennung der weiteren Faktoren nach HOFER (1969, S. 60) erfolgen.

Der Reliabilitätskoeffizient der Dimension Empfindsamkeit vs. Robustheit fällt gegenüber den anderen Dimensionen auf 0,52 ab, da dieser "Subtest" nur aus zwei Items (Polaritäten) besteht.

4.3.2. Clusteranalytische Ermittlung von Urteilstendenzen bei den einzelnen Lehrern

4.3.2.1. Klassifikation ähnlich beurteilter Schüler mit der Hierarchical Grouping Analysis nach WARD

Frage 1, ob sich bei den einzelnen Lehrern Beurteilertendenzen nachweisen lassen, bearbeiten wir mit der Clusteranalyse als Methode statistischer Typenfindung. Wir suchen in einzelnen Klassen nach unterscheidbaren Schülergruppen, die aufgrund ihrer Ähnlichkeit bezüg-

Tabelle 4b

Ergebnisse der Faktoren- und Itemanalyse der polaren Eigenschaften zur aktuellen Schülerbeurteilung

		Faktorenladungen						
I.	Disziplniertheit vs. Undisziplniertheit	I	II	III	IV	V	r_{it}	$\bar{x}$
	1. uninteressiert/interessiert	-.56	-.27	-.51	-.20	.03	.71	3.02
	9. fleißig/faul	.83	-.06	.14	.22	.03	.85	3.22
	12. ehrgeizig/gleichgültig	.83	.17	.21	.09	.14	.81	3.63
	13. unkonzentriert/konzentriert	-.71	.24	-.33	-.05	.24	.81	3.80
	15. unordentlich/ordentlich	-.60	.23	-.13	-.27	.08	.71	3.21
	19. pflichtbewußt/ pflichtvergessen	.83	-.12	.11	.24	-.02	.85	3.11
	20. aufmerksam/abgelenkt	.75	.18	.32	.15	-.18	.85	3.55
	23. zuverlässig/ unzuverlässig	.72	-.06	.16	.41	-.10	.83	2.94
II.	Zurückhaltung vs Geltungsstreben							
	2. verschlossen/ offen	-.19	-.72	-.15	-.13	-.28	.72	3.46
	4. selbstsicher/ selbstunsicher	.05	.49	.42	-.01	-.47	.67	3.81
	7. gesellig/ ungesellig	-.03	.68	.21	.33	-.22	.68	2.97
	8. geltungsbedürftig/ bescheiden	-.25	.68	.04	-.38	.07	.74	4.19
	16. schüchtern/ dreist	-.24	-.71	-.09	.41	.12	.79	4.17
	18. unruhig/ ruhig	-.39	.73	.03	-.15	.15	.74	4.62
	25. führend/ zurückhaltend	-.04	.75	.20	-.15	-.18	.82	4.19
III.	Geistige Regsamkeit							
	5. unbegabt/ begabt	-.23	-.01	-.89	-.08	.04	.88	3.41
	6. einfallsreich/ einfallslos	.22	.33	.71	.05	-.05	.83	3.64
	11. intelligent/ dumm	.21	.07	.87	.08	.05	.88	3.40
	21. aufgeweckt/träge	.34	.43	.62	.09	-.07	.83	3.37
IV.	Normenkonformität							
	14. höflich/ frech	.31	-.19	.06	.72	.01	.79	2.58
	17. folgsam/ unfolgsam	.40	-.33	-.03	.65	-.11	.84	2.84
	22. ausgeglichen/ launisch	.33	-.28	.09	.53	-.44	.80	3.29
	24. sympathisch/ unsympathisch	.21	.21	.24	.70	.00	.67	2.65
V.	Empfindsamkeit vs. Robustheit							
	3. sensibel/ robust	.17	-.44	.00	.26	.65	.83	3.69
	10. kompliziert/ unkompliziert	-.20	-.18	.05	-.24	.74	.81	4.14

Die Reliabilitätsschätzungen nach KUDER RICHARDSON 21 für die einzelnen als Subtests betrachteten Dimensionen fallen wie folgt aus:

I.	Disziplniertheit vs. Undisziplniertheit	.92
II.	Zurückhaltung vs. Geltungsstreben	.86
III.	Geistige Regsamkeit	.87
IV.	Normenkonformität	.78
V.	Empfindsamkeit vs. Robustheit	.52

lich ihrer vom Lehrer erhaltenen Beurteilungsscores auf den fünf Urteilsdimensionen zusammengefaßt werden können. Auch die Faktorenanalyse stellt eine Methode zur statistischen Typenfindung dar. Während in der Clusteranalyse Gruppen ähnlicher Meßwertträger gesucht werden, arbeitet die Faktorenanalyse (Q-Analyse) zugrundeliegende Dimensionen (Typusfaktoren) heraus. Dabei können zum Teil unähnliche Meßwertträger (Elemente) im selben Faktor laden. Allerdings sagt gleiche Ladungshöhe im selben Faktor noch nichts über die Ähnlichkeit der Meßwertträger aus, weil Unterschiede hinsichtlich der Lage (evaluation) zweier Profile und die Varianzunterschiede zweier Profile (scatter) nicht in die Faktorenanalyse eingehen. Ein Verfahren, welches diese beiden Profilaspekte berücksichtigt, findet sich bei VELDMAN (1968, S. 308) in der Hierarchical Grouping Analysis nach WARD (1963), die mit dem Programm HGROUP automatisiert worden ist. Das Verfahren betrachtet zunächst die N Profile (bezüglich k Variabler) der zu Gruppen zusammenzufassenden Meßwertträger als N Gruppen. Die beiden ähnlichsten dieser N Gruppen werden zunächst zusammengefaßt, es resultieren N minus 1 Gruppen. Das Kriterium für die Zusammenfassung zweier dieser N Gruppen bildet die relativierte Quadratsumme der Abweichungen dieser beiden Gruppen von ihrem gemeinsamen Mittelwertsprofil. Dieses Kriterium zeigt gleichzeitig das Ausmaß des Fehlerzuwachses an, den man bei iterierender Zusammenfassung von Gruppen begeht. Nun wird bei den N-1 Gruppen dasselbe Verfahren wiederholt. Als Abbruchkriterium dient hierbei die Inspektion des Fehlerzuwachses nach jeder Iteration. Man entscheidet sich für diejenige Gruppierung, die - verglichen mit vorherigen Iterationen - vor einem sprunghaften Fehlerzuwachs liegt.

Mit der Hierarchical Grouping Analysis ermitteln wir, ob sich pro Klasse Gruppen vom jeweiligen Lehrer unterschiedlich beurteilter Schüler finden lassen. Da 36 Leh-

rer die Schüler ihrer jeweiligen Klasse beurteilt hatten, sind 36 Clusteranalysen zu rechnen, wobei in jeder der 36 Clusteranalysen von unterschiedlich vielen Einzelprofilen ausgegangen werden muß, da die Klassen unterschiedlich viele Schüler (Meßwertträger) haben, wie es Tabelle 3 unter anderem ausweist.

Nach Inspektion der Rechenprotokolle der Clusteranalysen (vergl. Tab. 1 bis 36 im Anhang) können wir uns, unter Zugrundelegung des Kriteriums des sprunghaften Fehleranstiegs, in allen Klassen für eine Dreiergruppierung entscheiden. Diese konsistente Dreiergruppierung überrascht, wenn man daran denkt, daß wir eine sehr heterogene Lehrerstichprobe mit unterschiedlich vielen Schülern pro Klasse in die Untersuchung einbezogen haben. Diejenigen Auszüge aus den Rechenprotokollen, aus denen für jede einzelne Klasse d i e Gruppierung erkennbar wird, die vor einem sprunghaften Fehleranstieg liegt, sind im Anhang in den Tabellen 1 bis 36 zu finden. Die Tabellen 1 bis 36 im Anhang enthalten, geordnet nach den clusteranalytisch ermittelten Schülergruppierungen pro Klasse, die Beurteilungsscores eines jeden Schülers auf den fünf Beurteilungsdimensionen, zudem die Gruppenmittelwerte auf den fünf Beurteilungsdimensionen. Den Beurteilungsscore pro Schüler auf einer Dimension erhalten wir, indem die Summe aus allen zu einer Dimension gehörenden polaren Eigenschaften (Items) für jeden Schüler einer Klasse gebildet wird.

Um zu ermitteln, welche und wieviele Urteilsdimensionen sich für die jeweilige Dreiergruppierung in den einzelnen Klassen als trennwirksam erwiesen haben (vergl. Fragestellung 2), überprüfen wir die Mittelwertsunterschiede zwischen den jeweils drei Gruppen pro Klasse bezüglich jeder Beurteilungsdimension auf Signifikanz. Die multiplen Mittelwertsvergleiche werden mit Hilfe des

SCHEFFE-Tests, unter Zugrundelegung eines Signifikanzniveaus von $p \leq 0{,}05$, vorgenommen. Nach SCHEFFE (s. SACHS 1969, S. 512) gilt für ungleich große und gleich große Stichprobengruppen

$$D_I = \sqrt{s^2_{in} \; (1/n_i + 1/n_j) \; (k-1) \; F(k-1; \nu_{s^2_{in}}; \alpha)}$$

s^2_{in} = Mittleres Quadrat des Versuchsfehlers

n_i, n_j = Stichprobenumfänge der verglichenen Mittelwerte

$\nu_{s^2_{in}}$ = Anzahl der Freiheitsgrade für s^2_{in}

Für jede Klasse mußten 5, d.h. insgesamt 5x36=180 SCHEFFE-Tests gerechnet werden. Die multiplen Vergleiche pro Klasse bezüglich jeder Urteilsdimension sind den Tabellen 1 bis 36 im Anhang beigefügt. Die SCHEFFE-Tests wurden auf einer PDP 8I mit einem Programm von HIMMELMANN (1971), das Varianzhomogenität und Abweichungen von der Normalverteilung mit berücksichtigt, gerechnet.

4.3.2.2. Profildarstellung der Ergebnisse der Hierarchical Grouping Analysis

4.3.2.2.1. Kriterien zur Profilinterpretation

Die Darstellung der Ergebnisse der Clusteranalysen soll, um dem einzelnen Lehrer eine anschauliche Rückmeldung bezüglich seiner Urteilstendenzen zu ermöglichen, zusätzlich graphisch erfolgen. Wir wählen (entsprechend 36 Lehrern) 36 Profildarstellungen (s. Abb. 2.1 bis 2.36), um die bei jedem einzelnen Lehrer ermittelten Schülergruppierungen durch sog. Gruppenprofile übersichtlicher zu gestalten. Da sich für alle urteilenden Lehrer aufgrund des Kriteriums des sprunghaften Fehleranstiegs

Abbildung 2. 1. - 2. 36

Graphische Darstellung der Gruppenprofile pro Klasse*)

Legende

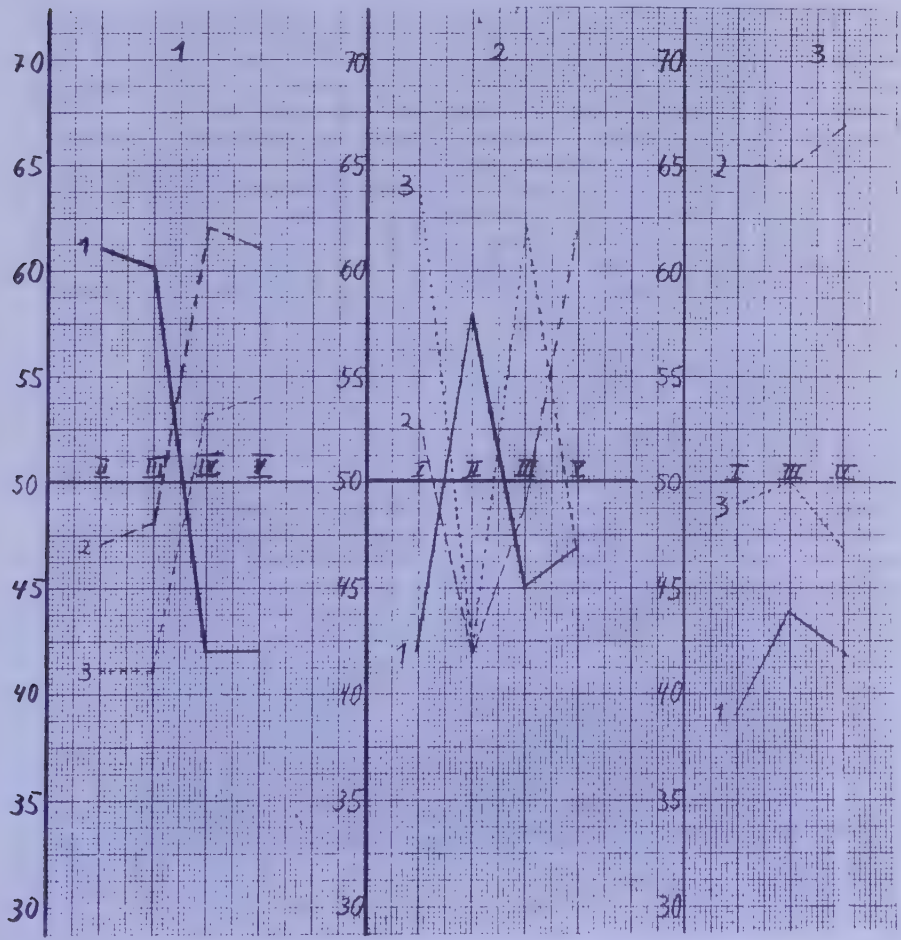
I : Disziplinierthheit vs. Undisziplinliertheit; kleine T-Werte bedeuten hohe Ausprägung im Sinne von disziplinliert

II : Zurückhaltung vs. Geltungsstreben; kleine T-Werte bedeuten hohe Ausprägung im Sinne von Geltungsstreben

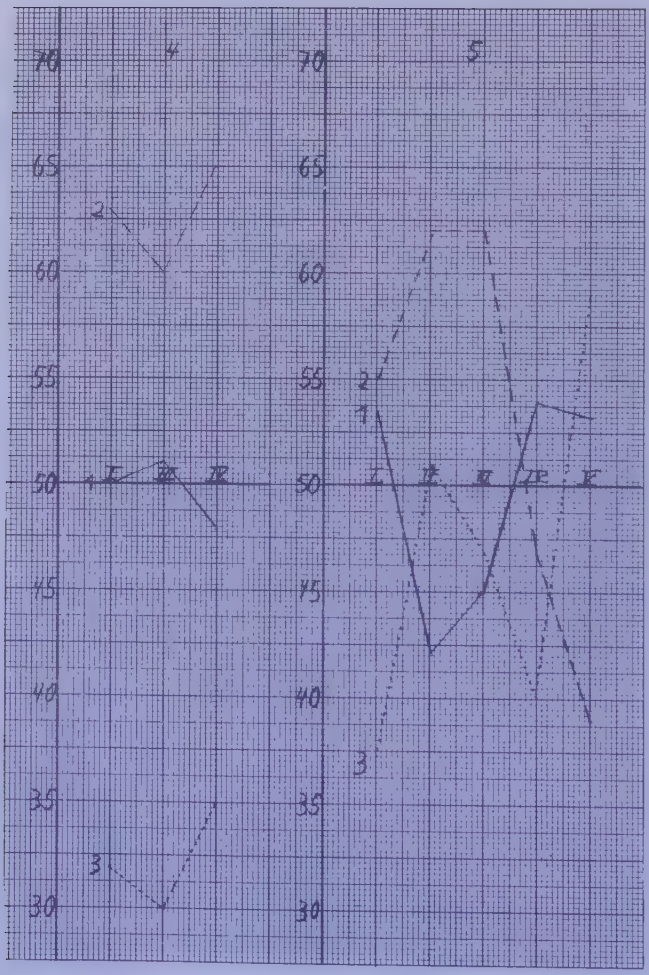
III : Geistige Regsamkeit; kleine T-Werte bedeuten hohe Ausprägung im Sinne von Begabung

IV : Normenkonformität; kleine T-Werte bedeuten hohe Ausprägung im Sinne von normenkonform

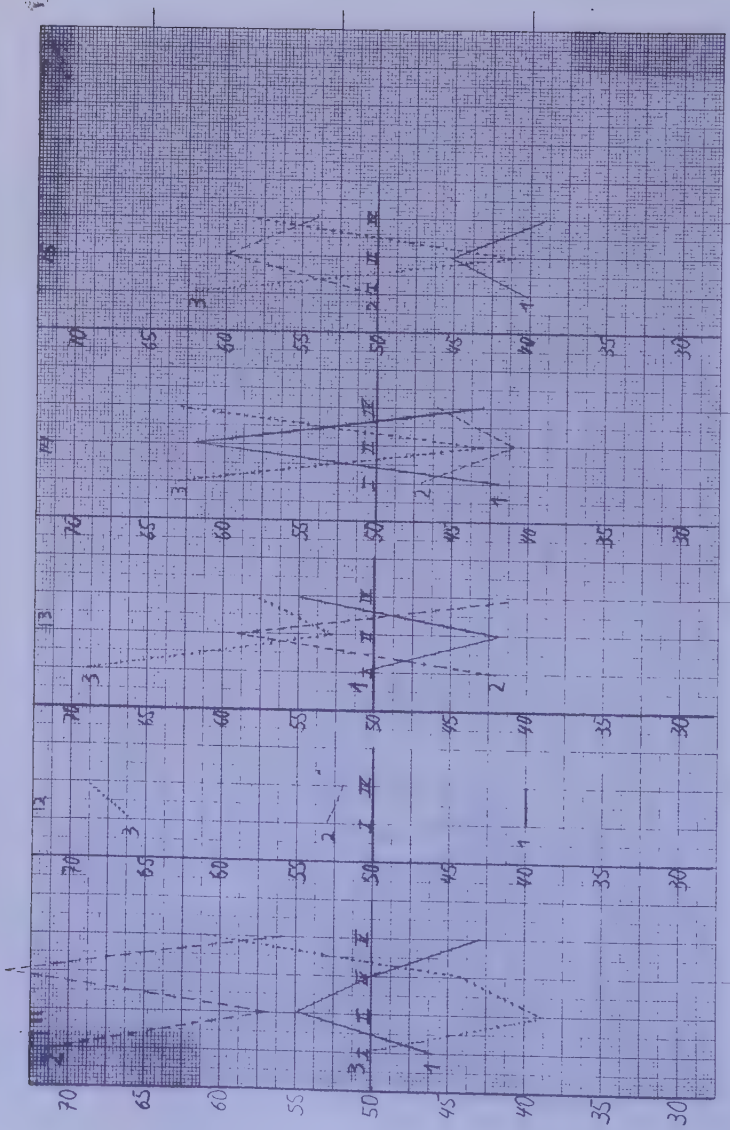
V : Empfindsamkeit vs. Robustheit; kleine T-Werte bedeuten hohe Ausprägung im Sinne von empfindsam

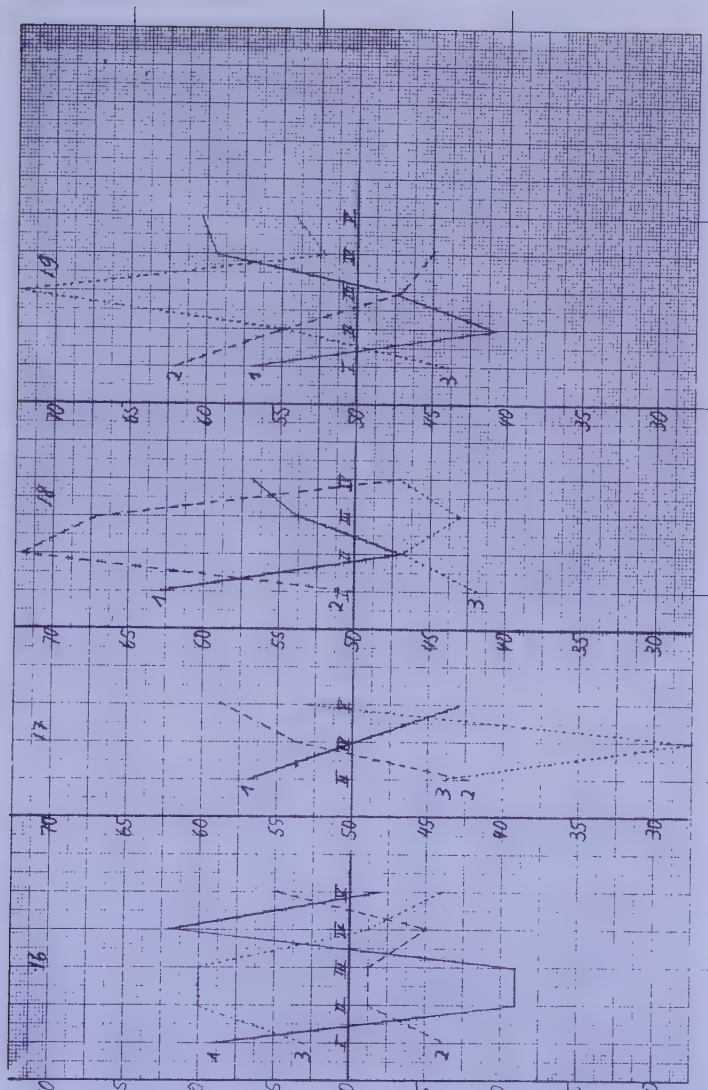


*Die drei Profile je Klasse sind zur besseren Unterscheidung durch unterschiedliche Verbindungslinien (Strichelung) gekennzeichnet. Interpretative Hinweise sind daraus nicht ableitbar.

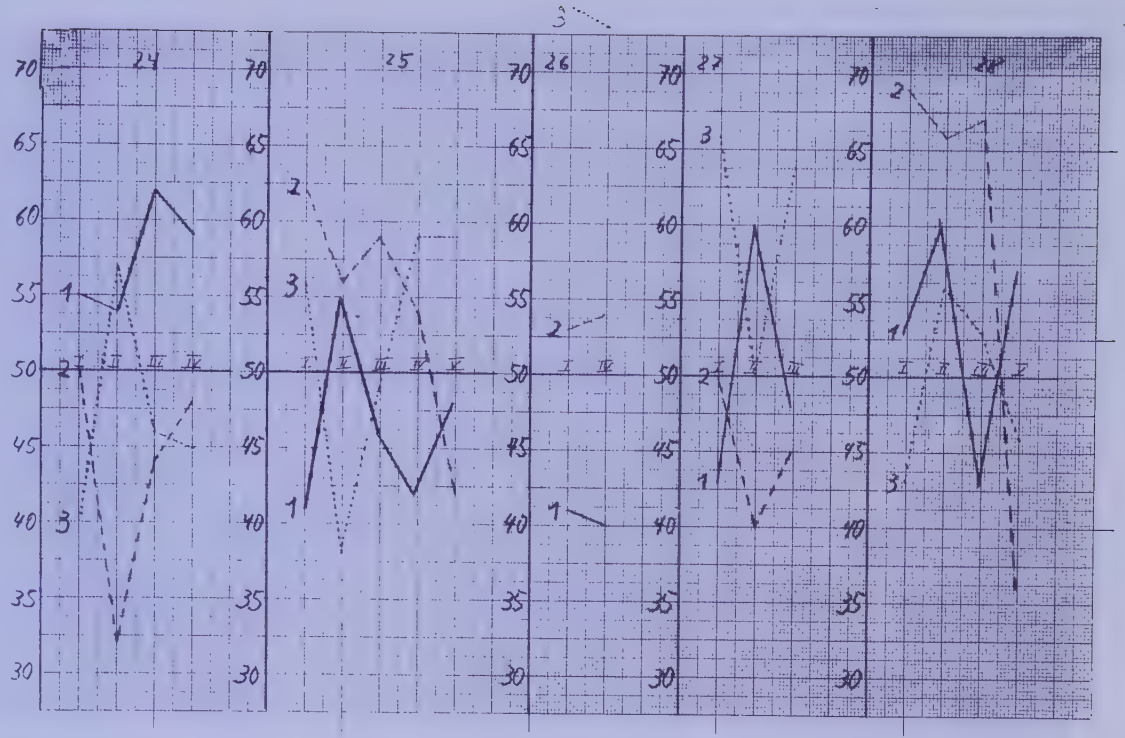

4
5
70
65
60
55
50
45
40
35
30
I
II
III
IV
V
1
2
3

70
65
60
55
50
45
40
35
30
6
7
8
9
10
1
2
3

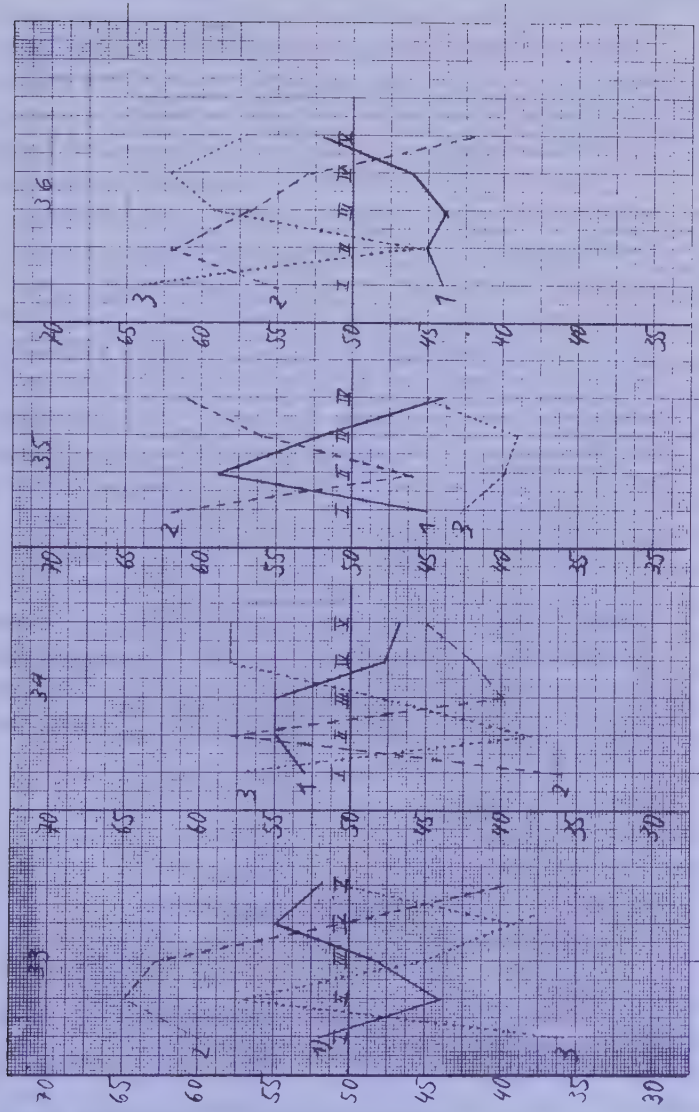

70
65
60
55
50
45
40
35
30
I
II
III
1
2
3


16
17
18
19
70
65
60
55
50
45
40
35
30
I
II
III
IV
1
2
3
4

70
65
60
55
50
45
40
35
30
I
II
III
IV
V
VI
1
2
3
21
22
23
30


24
25
26
27
28
70
65
60
55
50
45
40
35
30
I
II
III
IV
V

24
30
37
42
70
65
60
55
50
45
40
35
30
I
II
III
IV
1
2
3


70
65
60
55
50
45
40
35
30
I
II
III
IV
V
VI
1
2
3

eine konsistent interpretierbare Dreiergruppierung anbot (vergl. die Tabellen 1 bis 36 im Anhang), sind pro Lehrer drei Gruppenprofile zu zeichnen, wobei ein Profil den Verlauf der Gruppenmittelwerte auf den signifikant trennenden Dimensionen darstellt, die für jeden Lehrer auf der Abszisse abgetragen sind. Um die Gruppenmittelwerte zwischen den bei jedem Lehrer infragekommenden Urteilsdimensionen vergleichen zu können, wurden die Gruppenmittelwerte in T-Werte transformiert und auf der Ordinate abgetragen. Zur interpretativen Beschreibung der Gruppenprofile halten wir nach formalen Kriterien Ausschau, um Fehl- und/oder Überinterpretationen weitgehend zu reduzieren. Ein formales Kriterium als Hilfsmittel zur Interpretation der Gruppenprofile ist bereits durch die SCHEFFE-Methode gegeben. Mit ihr wurden die Urteilsdimensionen pro Lehrer ermittelt, die signifikant zur Trennung der Schülergruppen beitragen. Eine Urteilsdimension wird nur dann berücksichtigt und auf der Abszisse abgetragen, wenn sie mindestens eine signifikante Differenz von drei möglichen ausweist. Als weiteres formales Kriterium bietet sich, wie bei Profilinterpretationen üblich, die signifikante Abweichung eines Gruppenmittelwertes vom Klassenmittelwert an. Diese Abweichung stellen wir mit Hilfe des Vertrauensintervalls (CL) fest, das über den Standardmeßfehler ($s_e = s_x \sqrt{1-r_{tt}}$) wie folgt definiert ist:

$$CL = \bar{X}_{(T)} \pm z \cdot s_x \sqrt{1-r_{tt}}$$

CL = Vertrauensintervall

$\bar{X}_{(T)}$ = Mittelwert der T-Verteilung ($\bar{X}_{(T)} = 50$)

z_α = Vorgegebene Irrtumswahrscheinlichkeit

s_x = Standardabweichung der T-Verteilung ($s_x = 10$)

r_{tt} = Zuverlässigkeit der jeweiligen Urteilsdimension, ermittelt pro Lehrer.

Die für die Berechnung des Vertrauensintervalls notwendige Zuverlässigkeit r_{stab} jeder einzelnen Dimension pro Lehrer ist in den Tabellen 1 bis 36 im Anhang mit aufgeführt.

Gruppenmittelwerte, die nicht signifikant vom Klassenmittelwert abweichen, können somit im Sinne des "Durchschnitts" interpretiert werden.

Als inhaltliche Interpretationshilfe zur Charakterisierung der einzelnen Profile bot sich an, in Anlehnung an die BIERIsche Version (BIERI et al., 1966) des "Role Construct Repertory Tests" (REP-Test), nach einer übergeordneten Bewertungsdimension zu suchen, d.h. aus den 25 vorgegebenen polaren Eigenschaftspaaren diejenigen auszuwählen, in denen neben einer Beschreibung spezieller Eigenschaften (z.B. "aufmerksam-abgelenkt", "interessiert -uninteressiert") eine bewertende Komponente im Sinne von positiv-negativ deutlich wird. Ein Skalierungsversuch mit 66 Lehrern erbrachte Aufschluß darüber, bei welchen Eigenschaftspaaren innerhalb der Lehrerstichprobe ein hohes Maß an Übereinstimmung hinsichtlich der Bipolarität des Kontinuums und hinsichtlich der Definition der Pole als negative bzw. positive bestand. Von den 12 Eigenschaftspaaren, die die genannten Kriterien am besten erfüllen, entfallen die überwiegende Anzahl (neun!) auf die Dimensionen I und III ("Disziplintiertheit vs. Undisziplintiertheit" und "geistige Regsamkeit"). Die Durchführung und Auswertung des Skalierungsversuchs erfolgte nach der Methode der sukzessiven Intervalle in Verbindung mit dem Gesetz vom kategorialen Urteil nach TORGERSON (1958, S. 205-246).

Um diejenigen der 25 Eigenschaftspaare herauszufinden, in denen neben einer Beschreibung spezieller Eigenschaften eine besonders bewertende Komponente zum Ausdruck kommt, entschieden wir uns, da das Verfahren des Paarvergleichs

zu aufwendig ist, für die Skalierungsmethode der sukzessiven Intervalle in Verbindung mit dem Gesetz vom kategorialen Urteil nach TORGERSON (1958, S. 205-246). Im folgenden werden die für unser praktisches Problem notwendigen Schritte der Erhebung und rechnerischen Auswertung dargelegt. Es werden zwei Skalierungsversuche mit je 33 Lehrern durchgeführt. Von den 25 polaren Eigenschaftspaaren wurde je eine bei zwei polaren Eigenschaften dem ersten, die entgegengesetzte Eigenschaft dem zweiten Skalierungsversuch derart zugeordnet, daß auf beide Versuche etwa gleich viele positive wie negative Eigenschaften entfielen. Für die beiden Versuche wurden die je 25 Eigenschaften einzeln auf Kärtchen geschrieben, die die Lehrer in eine Rangreihe von fünf Haufen zu ordnen hatten (Sortiermethode). Die Instruktion für die Lehrer lautete:

"Ich lege Ihnen 25 Kärtchen vor, auf denen Eigenschaften stehen, mit denen Lehrer Schüler beurteilen. Bitte verteilen Sie die Karten auf fünf Haufen, wobei jeder Haufen einen Ausprägungsgrad darstellt. Ordnen Sie die Karten danach für wie positiv bzw. negativ Sie die betreffende Eigenschaft bei Schülern halten.
Auf den ersten Haufen legen Sie bitte die Eigenschaften, die Ihrer Meinung nach positiv für einen Schüler sind, auf den letzten Haufen nur Eigenschaften, die negativ für einen Schüler sind. Die restlichen drei Haufen stellen Abstufungen zwischen den Extremen dar."

Die analytische Auswertung der Raterurteile erfolgte nach TORGERSON (1960, S. 211 - 213) in sechs Schritten, wobei jeder Schritt durch eine Matrix gekennzeichnet ist:

1. Schritt: F-Matrix:

In diese Matrix werden die Ergebnisse der Raterurteile eingetragen. Die Zellen dieser Matrix geben die Häufigkeit an, wie oft Item j in die Kategorie g eingestuft wurde. Es gibt eine Kategoriengrenze weniger als es

Kategorien gibt, da die erste und letzte Kategorie offenstehen. Weil die Anzahl der Kategoriengrenzen gleich m ist, gibt es also m+1 Kategorien.

2. Schritt: $\int\!\!\int$ - Matrix:

Die $\int\!\!\int$ - Matrix enthält die von links aufkumulierten Häufigkeiten der Matrix F.

3. Schritt: P - Matrix:

Die P-Matrix enthält die auf 1 relativierten Werte der $\int\!\!\int$ -Matrix (Werte der $\int\!\!\int$ -Matrix durch N dividiert). Die letzte Spalte (m+1) entfällt, da in dieser Spalte überall die Zahl 1 steht.

4. Schritt: X - Matrix:

Die X-Matrix ergibt sich aus der P-Matrix durch z-Transformation. Folgendes ist zu beachten: Tritt in der X-Matrix 0 oder 1 auf, so erhält man durch Transformation für $z = \pm\infty$. Da man mit unendlichen Werten nicht weiterrechnen kann, läßt man diese Felder leer; sie sind durch einen Querstrich gekennzeichnet.

5. Schritt: ΔX- Matrix (Differenzenmatrix):

Durch die Differenzenmatrix werden die Kategoriengrenzen ermittelt. Hierbei ist die Differenz nebeneinanderstehender z-Werte gleich der Differenz zwischen den entsprechenden Kategoriengrenzen. Wegen der Meßfehler erhält man jedoch nur Schätzwerte der Differenz $t_{g+1} - t_g$. Die beste Schätzung der Differenz ist der Mittelwert über alle Stimuli j. Wir haben somit eine Möglichkeit, den Abstand zwischen aufeinanderfolgenden Kategoriengrenzen zu schätzen. Zu den Skalenwerten gelangt man, indem man zunächst die erste Kategorien-

grenze $t_1=0$ setzt. Dann liegen die anderen Kategoriengrenzen tg ebenfalls fest, da die Differenzen bekannt sind.

6. Schritt: $((tg - \Sigma jg))$ - Matrix:

Die Skalenwerte Sj der Stimuli erhält man durch Bildung der Differenzen tg - zjg für jede Kategorie; die beste Schätzung für Sj ist der Mittelwert der Differenzen.

Nach der Beschreibung der Rechenschritte soll an dieser Stelle die Verrechnung des Zahlenmaterials des Expertenratings folgen. Diejenigen 12 Eigenschaftspaare (ca. die Hälfte von 25), in denen am stärksten eine bewertende Komponente zum Ausdruck kommt, wurden durch die Summe der beiden absoluten Differenzen von den Skalenmittelwerten ermittelt. Ausgewählt wurden die 12 polaren Eigenschaftspaare mit der größten Summe. Sie sind in Tab.5 mit einem Stern versehen.

Bevor wir uns der graphischen Darstellung (s. Abb. 2.1. bis 2.36, S.75/76) und der Interpretation einzelner Dreierprofile, entsprechend der Dreiergruppierung pro Lehrer, zuwenden, möchten wir darauf hinweisen, warum es nicht Sinn dieser Arbeit sein kann, die an der Untersuchung beteiligten Lehrer danach zu klassifizieren, wie ähnlich ihre Dreierprofile sind (vergl. dazu die Dreierprofilverläufe der Lehrer 3, 4, 12 und 26). Da wir mit dieser Arbeit ein Instrumentarium entwickeln und erproben wollen, um sehr verschiedenen Lehrern Rückmeldung über ihr realisiertes Urteilsverhalten zu verschaffen, wurde eine möglichst heterogene Lehrerstichprobe gewählt. So wurden in die Stichprobe Lehrer unterschiedlichen Alters, Geschlechts, Schultyps sowie unterschiedlicher Ausbildung aufgenommen. Zudem streute die Klassenstärke zwischen 10 und 37 Schülern beim einzelnen Lehrer.

Tab. 5

Auswertung der Daten des Expertenratings

Item Nr.		F - Matrix Kategorien					∮ - Matrix Kategorien					P - Matrix Kategorien				X - Matrix Kategorien			
		-2	-1	0	+1	+2	-2	-1	0	+1	+2	-2	-1	0	+1	-2	-1	0	+1
1	uninteressiert	21	10	2	-	-	21	31	33	33	33	0.64	0.94	1.00	1.00	0.36	1.55	-	-
2	verschlossen	8	18	7	-	-	8	26	33	33	33	0.24	0.79	1.00	1.00	-0.70	0.81	-	-
3	sensibel	-	3	17	7	6	-	3	20	27	33	-	0.09	0.61	0.82	-	-1.34	0.28	0.91
4	selbstsicher	-	1	4	15	13	-	1	5	20	33	-	0.03	0.15	0.61	-	-1.88	-1.04	0.28
5	unbegabt	10	17	4	2	-	10	27	31	33	33	0.30	0.82	0.94	1.00	-0.52	0.91	1.55	-
6	einfallsreich	-	-	-	3	30	-	-	-	3	33	-	-	-	0.09	-	-	-	-1.34
7	gesellig	-	1	4	12	16	-	1	5	17	33	-	0.03	0.15	0.51	-	-1.88	-1.04	0.03
8	geltungsbedürftig	15	7	-	-		11	26	33	33	33	0.33	0.79	1.00	1.00	-0.44	0.81	-	-
9	fleißig	-	-	2	21	10	-	-	2	23	33	-	-	0.06	0.70	-	-	-1.55	0.52
10	kompliziert	2	15	8	6	2	2	17	25	31	33	0.06	0.51	0.76	0.94	-1.55	0.03	0.71	1.55
11	intelligent	-	-	3	5	25	-	-	3	8	33	-	-	0.09	0.24	-	-	-1.34	-0.70
12	ehrgeizig	-	6	6	16	5	-	6	12	28	33	-	0.18	0.36	0.85	-	-0.92	-0.35	1.04
13	unkonzentriert	15	16	2	-	-	15	31	33	33	33	0.45	0.94	1.00	1.00	-0.13	1.55	-	-
14	höflich	1	-	15	13	4	1	1	16	29	33	0.03	0.03	0.49	0.88	-1.88	-1.88	-0.03	1.17
15	unordentlich	13	14	6	-	-	13	27	33	33	33	0.39	0.82	1.00	1.00	-0.28	0.91	-	-
16	schüchtern	-	10	22	1	-	-	10	32	33	33	-	0.30	0.97	1.00	-	-0.52	1.88	-
17	folgsam	2	1	14	14	2	2	3	17	31	33	0.06	0.09	0.51	0.94	-1.55	-1.34	0.03	1.55
18	unruhig	6	20	7	-	-	6	26	33	33	33	0.18	0.79	1.00	1.00	-0.92	0.81	-	-
19	pflichtbewußt	-	-	4	21	8	-	-	4	25	33	-	-	0.12	0.76	-	-	-1.17	0.71
20	empfindsam	-	1	-	15	17	-	1	1	16	33	-	0.03	0.03	0.49	-	-1.88	-1.88	-0.03
21	aufgeweckt	-	-	-	11	21	-	-	1	12	33	-	-	0.03	0.36	-	-	-1.88	-0.35
22	ausgeglichen	-	-	2	16	15	-	-	2	18	33	-	-	0.06	0.54	-	-	-1.55	0.10
23	zuverlässig	2	-	2	11	18	2	2	4	15	33	0.06	0.06	0.12	0.45	-1.55	-1.55	-1.17	-0.13
24	sympathisch	-	-	12	9	12	-	-	12	21	33	-	-	0.36	0.64	-	-	-0.35	0.36
25	führend	-	4	13	14	2	-	4	17	31	33	-	0.12	0.51	0.94	-	-1.17	0.03	1.55
1	interessiert	-	-	-	2	31	-	-	-	2	33	-	-	0.06	0.06	-	-	-1.55	-1.55
2	offen	-	1	-	3	29	-	1	1	4	33	-	0.03	0.03	0.12	-	-1.88	-1.88	-1.17
3	robust	1	-	14	14	4	1	1	15	29	33	0.03	0.03	0.45	0.88	-1.88	-1.88	-0.13	1.17
4	selbstsicher	4	16	13	-	-	4	20	33	33	33	0.12	0.61	1.00	1.00	-1.17	0.28	-	-
5	begabt	-	-	-	8	25	-	-	-	8	33	-	-	-	0.24	-	-	-	-0.70
6	einfallslos	13	14	6	-	-	13	27	33	33	33	0.39	0.82	1.00	1.00	-0.28	0.91	-	-
7	ungesellig	6	16	8	3	-	6	22	30	33	33	0.18	0.67	0.91	1.00	-0.92	0.44	1.34	-
8	bescheiden	-	6	9	15	3	-	6	15	30	33	-	0.18	0.45	0.91	-	-0.92	-0.13	-
9	faul	17	9	7	-	-	17	26	33	33	33	0.51	0.79	1.00	1.00	0.03	0.81	-	-
10	unkompliziert	-	-	9	13	11	-	-	9	22	33	-	-	0.27	0.67	-	-	-0.61	0.44
11	dumm	12	5	15	1	-	12	17	32	33	33	0.36	0.51	0.97	1.00	-0.35	0.03	1.88	-
12	gleichgültig	11	11	9	2	-	11	22	31	33	33	0.33	0.67	0.94	1.00	-0.44	0.44	1.55	-
13	konzentriert	-	-	-	5	28	-	-	-	5	33	-	-	-	0.15	-	-	-	-1.04
14	frech	5	17	6	5	-	5	22	28	33	33	0.15	0.67	0.85	1.00	-1.04	0.44	1.04	-
15	ordentlich	-	2	8	17	6	-	2	10	27	33	-	0.06	0.30	0.82	-	-1.55	-0.52	0.91
16	dreist	10	11	7	4	1	10	21	28	32	33	0.30	0.64	0.85	0.97	-0.52	0.36	1.04	1.88
17	unfolgsam	5	12	11	4	1	5	17	28	32	33	0.15	0.51	0.85	0.97	-1.04	0.03	1.04	1.88
18	ruhig	1	2	8	19	3	1	3	11	30	33	0.03	0.09	0.33	0.91	-1.88	-1.34	-0.44	1.34
19	pflichtvergessen	12	14	6	1	-	12	26	32	33	33	0.36	0.79	0.97	1.00	-0.35	0.81	1.88	-
20	abgelenkt	3	21	8	1	-	3	24	32	33	33	0.09	0.72	0.97	1.00	-1.34	0.58	1.88	-
21	träge	10	20	3	-	-	10	30	33	33	33	0.30	0.91	1.00	1.00	-0.52	1.34	-	-
22	launisch	8	15	10	-	-	8	23	33	33	33	0.24	0.70	1.00	1.00	-0.70	0.52	-	-
23	unzuverlässig	21	10	1	1	-	21	31	32	33	33	0.64	0.94	0.97	1.00	0.36	1.55	1.88	-
24	unsympathisch	8	3	22	-	-	8	11	33	33	33	0.24	0.33	1.00	1.00	-0.70	-0.44	-	-
25	zurückhaltend	1	4	11	17	-	1	5	16	33	33	0.03	0.15	0.49	1.00	-1.88	-1.04	-0.03	-

Fortsetzung von Tab. 5

Item Nr.	Δ X-Matrix				$((t_g - z_{jg}))$ - Matrix					
	2 - 1	3 - 2	4 - 3	$t_1 - z_{j1}$	$t_2 - z_{j2}$	$t_3 - z_{j3}$	$t_4 - z_{j4}$	Summe	Skalenwerte	$\lvert\bar{X}_1 - X_j\rvert + \lvert\bar{X}_2 - X_j\rvert$
1	1.19	-	-	-0.36	-0.52	-	-	-0.88	-0.44	5.55*
2	1.51	-	-	0.70	0.26	-	-	0.96	0.48	3.17*
3	-	1.62	0.63	-	2.41	1.82	2.47	6.70	2.23	0.37
4	-	0.84	1.32	-	2.95	3.14	3.10	9.19	3.06	1.42
5	1.43	0.64	-	0.52	0.16	0.55	-	1.23	0.41	3.81*
6	-	-	-	-	-	-	4.72	4.72	4.72	3.84*
7	-	0.84	1.07	-	2.95	3.14	3.35	9.44	3.15	1.76
8	1.25	-	-	0.44	0.26	-	-	0.70	0.35	2.37*
9	-	-	2.07	-	-	3.65	2.86	6.51	3.26	2.48*
10	1.58	0.68	0.81	1.55	1.04	1.39	1.83	5.81	1.45	1.71
11	-	-	0.64	-	-	3.44	4.08	7.52	3.76	2.60*
12	-	0.57	1.39	-	1.99	2.45	2.34	6.78	2.26	1.10
13	1.68	-	-	0.13	-0.48	-	-	-0.35	-0.18	4.74*
14	0.00	1.85	1.20	1.88	2.95	2.13	2.24	9.20	2.30	0.77
15	1.19	-	-	0.28	0.16	-	-	0.44	0.22	2.80*
16	-	2.40	-	-	1.59	0.22	-	1.81	0.60	2.11
17	0.21	1.31	1.52	1.55	2.41	2.07	1.86	7.89	1.97	0.53
18	1.73	-	-	0.92	0.26	-	-	1.18	0.59	2.64*
19	-	-	1.88	-	-	3.27	2.67	5.94	2.97	2.07
20	-	0.00	1.85	-	2.95	3.98	3.41	10.34	3.47	2.17
21	-	-	1.53	-	-	3.98	3.73	7.71	3.86	3.07*
22	-	-	1.70	-	-	3.65	3.23	6.88	3.44	2.15
23	0.00	0.38	1.04	1.55	2.62	3.27	3.51	10.95	2.74	2.39*
24	-	-	0.71	-	-	2.45	3.02	5.47	2.74	0.97
25	-	1.20	1.52	-	2.24	2.07	1.86	6.07	2.04	1.62
1	-	-	-	-	-	-	4.43	4.43	4.43	
2	-	0.00	0.71	-	0.92	3.87	4.05	8.84	2.95	
3	0.00	1.75	1.30	1.88	0.92	2.12	1.71	6.63	1.66	
4	1.45	-	-	1.17	0.82	-	-	1.99	1.00	
5	-	-	-	-	-	-	3.58	3.58	3.58	
6	1.19	-	-	0.28	0.19	-	-	0.47	0.24	
7	1.36	0.90	-	0.92	0.66	0.65	-	2.23	0.75	
8	-	0.79	-	-	2.02	2.12	-	4.14	2.08	
9	0.78	-	-	-0.03	0.29	-	-	0.26	0.13	
10	-	-	1.05	-	-	2.60	2.44	5.04	2.52	
11	0.38	1.85	-	0.35	1.07	0.11	-	1.53	0.51	
12	0.88	1.11	-	0.44	0.66	0.44	-	1.54	0.51	
13	-	-	-	-	-	-	3.92	3.92	3.92	
14	1.48	0.60	-	1.04	0.66	0.95	-	2.65	0.89	
15	-	1.03	1.43	-	2.65	2.51	1.97	7.13	2.38	
16	0.88	0.68	0.84	0.52	0.74	0.95	1.00	3.21	0.81	
17	1.07	1.01	0.84	1.04	1.07	0.95	1.00	4.06	1.02	
18	0.54	0.90	0.00	1.88	2.44	2.43	1.54	8.26	2.07	
19	1.16	1.07	-	0.35	0.29	0.11	-	0.75	0.25	
20	1.92	1.30	-	1.34	0.52	0.11	-	1.97	0.66	
21	1.86	-	-	0.52	-0.24	-	-	0.28	0.14	
22	1.22	-	-	0.70	0.58	-	-	1.28	0.64	
23	1.19	0.33	-	-0.36	-0.45	0.11	-	-0.84	-0.29	
24	0.26	-	-	0.70	1.54	-	-	2.24	1.12	
25	0.84	1.01	-	1.88	2.14	2.02	-	(.04	2.02	
Σ	18.46	14.33	5.33							
N	18	15	6							
$\bar{X}$	1.03	0.96	0.89							

Kategoriengrenzen(oberer Versuch):

t_1=0
t_2=1.07
t_3=2.10
t_4=3.41

Kategoriengrenzen(unterer Versuch):

t_1=0
t_2=1.03
t_3=1.99
t_4=2.88

4.3.2.2.2. Interpretationen einzelner Profilabbildungen

Neben den graphischen Darstellungen der Ergebnisse der 36 Clusteranalysen durch Profilabbildungen haben wir im vorangegangenen Kapitel die Kriterien zur Interpretation der einzelnen Profilabbildungen herausgestellt. Diese Kriterien seien an dieser Stelle kurz wiederholt. Es sind

(1) Die SCHEFFE - Methode zur Auswahl signifikant trennender Urteilsdimensionen;

(2) Die signifikante Abweichung eines Gruppenmittelwertes vom Klassenmittelwert, die durch das Vertrauensintervall (CL) über den Standardmeßfehler definiert ist.

(3) Die Dimensionen "Disziplinertheit vs. Undisziplinertheit" und "geistige Regsamkeit" als Bewertungskriterien im Sinne von positiv - negativ.

An dieser Stelle soll nun nicht die Interpretation aller 36 Profilabbildungen erfolgen. Um die Anwendung der Interpretationskriterien zu demonstrieren, greifen wir uns aus den 36 Profilabbildungen fünf solche mit sehr unterschiedlichen Dreierprofilverläufen heraus, die wir interpretativ beschreiben.

Wir wählen dazu die Profilabbildungen der Lehrer 4, 12, 18, 29 und 36 aus, die wir aus Gründen der Anschaulichkeit noch einmal aufzeichnen. Signifikante Abweichungen eines Gruppenmittelwertes vom Klassenmittelwert (T=50) sind umkreist. Kleine T-Werte bedeuten hohe Ausprägungsgrade auf der jeweiligen Beurteilungsdimension.

Abbildung 3.1 bis 3.5

Graphische Darstellung und interpretative Beschreibung der Ergebnisse von fünf ausgewählten Clusteranalysen (Lehrer 4, 12, 18, 29, 36).

Legende: I =Disziplniertheit vs. Undisziplniertheit (Arbeitsverhalten)
II =Soziale Aufgeschlossenheit vs. Verschlossenheit
III=Geistige Regsamkeit (Begabung)
IV =Normenkonformität
V =Empfindsamkeit vs. Robustheit

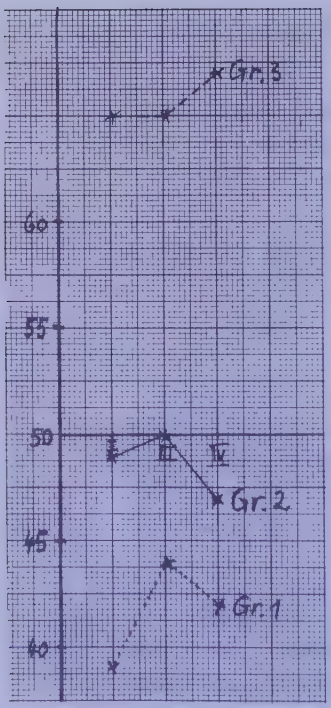


Abb. 3.1: Lehrer 4

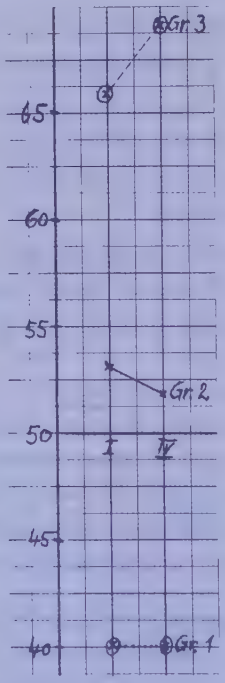


Abb. 3.2: Lehrer 12

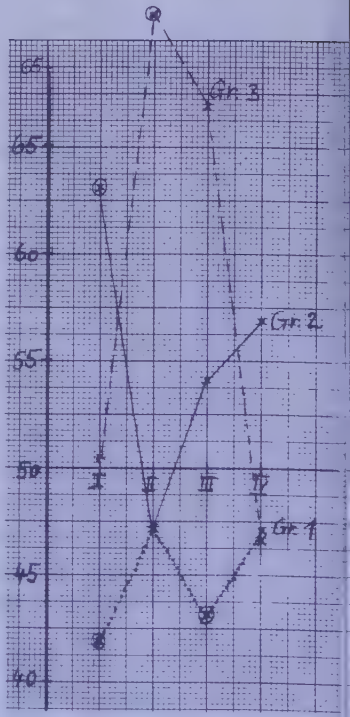


Abb. 3.3: Lehrer 18

Bei Lehrer 4 erfolgt die Gruppentrennung aufgrund der Dimensionen Disziplniertheit(I), geistige Regsamkeit(III) und Normenkonformität(IV).Gruppe 1 ist in der Sicht des Lehrers diszipliniert im Sinne guten Arbeitsverhaltens, begabt und normenkonform. Gruppe 3 ist demgegenüber gekennzeichnet durch mangelndes Arbeitsverhalten, geringe Begabung und wenig Normenkonformität. Gruppe 2 liegt bezüglich der genannten Dimensionen im Mittelbereich und stellt die größte Gruppe dar. Gruppe 1 stellt unter dem Bewertungsaspekt die positiv und Gruppe 3 die negativ eingestufte Gruppe dar.

Bei Lehrer 12 lassen sich nur zwei Dimensionen (I und IV) als trennwirksam nachweisen. Gruppe 1 zeichnet sich in in der Sicht des Lehrers durch gutes Arbeitsverhalten und hohe Normenkonformität aus, Gruppe 3 steht in deutlichem Gegensatz zu Gruppe 1. Gruppe2liegt bezüglich der genannten Dimensionen im Mittelbereich. Gruppe 1 stellt somit die positiv und Gruppe 3 die negativ bewertete Gruppe dar.

Bei Lehrer 18 erfolgt die Gruppentrennung aufgrund der Dimensionen Disziplniertheit (I), Zurückhaltung (II), geistige Regsamkeit (III) und Normenkonformität (IV), wobei bei Dimension IV auffällt, daß sich Gruppe 1 und 3 signifikant von Gruppe 2 unterscheiden, aber alle drei Gruppen im Durchschnittsbereich liegen. Gruppe 1 zeichnet sich durch überdurchschnittliche Werte in Bezug auf Disziplniertheit und Begabung aus. Gruppe 2 fällt als überdurchschnittlich undisziplniert auf. Gruppe 3 ist weit überdurchschnittlich zurückhaltend bei gleichzeitig mangelnder geistiger Regsamkeit. Gruppe 1 läßt sich eindeutig als die vom Lehrer positiv gesehene Gruppe interpretieren.

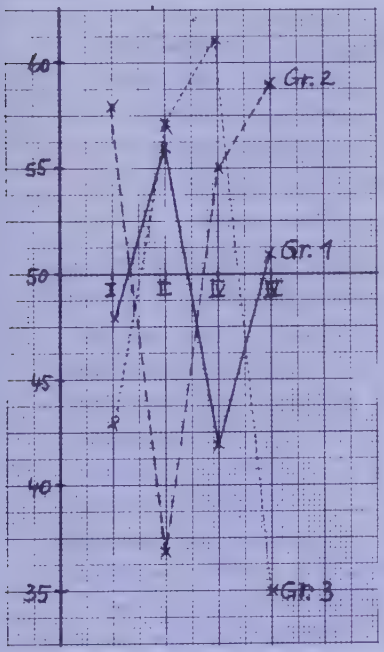


Abb. 3.4: Lehrer 29

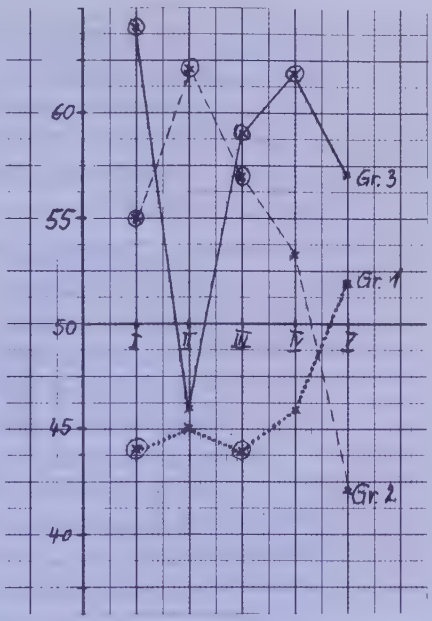


Abb. 3.5: Lehrer 36

Bei Lehrer 29 erfolgt die Gruppentrennung aufgrund der Dimensionen I,II,IV und V. Gruppe 3 zeichnet sich durch überdurchschnittliche Disziplniertheit im Sinne guten Arbeitsverhaltens, Zurückhaltung, mangelnder Normenkonformität und Empfindsamkeit aus. Gruppe 2 dagegen ist in der Sicht des Lehrers undisziplniert (im Sinne schlechten Arbeitsverhaltens), geltungsstrebig und robust. Gruppe 1 fällt nur durch Normenkonformität auf. Unter dem Bewertungsaspekt lassen sich diese Gruppen nicht eindeutig interpretieren, da auf der Dimension geistige Regsamkeit (III) keine signifikanten Gruppenunterschiede nachgewiesen werden konnten.

Bei Lehrer 36 erfolgt die Gruppentrennung aufgrund aller fünf Dimensionen, wobei bezüglich Dimension V keine signifikanten Mittelwertsabweichungen nachweisbar sind. Gruppe 1 zeigt in der Sicht des Lehrers ein überdurchschnittlich gutes Arbeitsverhalten und überdurchschnittlich hohe Intelligenz. Gruppe 2 ist bei mangelndem Arbeitsverhalten durch Zurückhaltung und geringe Begabung gekennzeichnet. Gruppe 3 ist bei mangelndem Arbeitsverhalten und geringer Intelligenz wenig normenkonform. Unter dem Bewertungsaspekt stellt Gruppe 1 eine große positiv beurteilte Gruppe dar (N=14), der zwei kleinere (N=9 und N=5) negativ bewertete Gruppen gegenüberstehen.

4.3.2.3. Abschließende Bemerkungen zur Interpretation einzelner Profile

Durch vorheriges Aufstellen der Interpretationskriterien (vergl. Kap. 4.3.2.2.1.) war die Interpretation einzelner Profile fest vorstrukturiert. Wir erhielten dadurch sehr vorsichtige und wenig spekulative, dafür aber statistisch abgesicherte Aussagen.

Dieses Vorgehen entspricht der Forderung der Interpretationsobjektivität (vergl. LIENERT 1967[2] S. 14), d.h. der Unabhängigkeit der Interpretation der Untersuchungsergebnisse von der Person des jeweiligen Autors. Dabei war es unser Anliegen, das Urteilsverhalten der einzelnen Lehrer zu objektivieren, d.h. dem einzelnen Lehrer vor Augen zu führen, wie sich aufgrund seiner aktuellen Schülereinschätzungen unterscheidbare Schülergruppen auffinden und charakterisieren lassen. Dabei haben wir uns einer Bewertung des Urteilsverhaltens im Sinne von "richtig" und "falsch" enthalten. Nicht die Frage nach der "Richtigkeit" (Validitätsaspekt) beurteilter Schülergruppen stand im Vordergrund, sondern vielmehr die Frage, wie sich der Prozeß der Urteilsbildung beim Lehrer abspielt.

Bei der interpretativen Beschreibung einzelner Profile haben wir die Faktorenbenennungen (s. Tab. 4b) verwandt. Der dadurch eintretende Informationsverlust läßt sich leicht beheben, indem man bei der Rücksprache mit einem Lehrer die zu einzelnen Dimensionen gehörenden polaren Eigenschaften mitteilt und dadurch die Rückmeldung konkretisiert.

4.3.3. Multiple Korrelationsanalysen

4.3.3.1. Bemerkungen zu den instrumentellen Eigenschaften der Prädiktor- und Kriteriumsvariablen

Die dritte Fragestellung, ob sich bei einzelnen Lehrern Zusammenhänge zwischen den fünf Urteilsdimensionen einerseits und bestimmten Noten andererseits aufzeigen lassen, spezifizieren wir dahingehend, ob sich die Zeugnisnoten als Kriteriumsvariablen aus den Dimensionen Disziplinviertheit vs. Undisziplinviertheit, Zurückhaltung vs. Geltungsstreben, geistige Regsamkeit, Normenkonformität und Empfindsamkeit vs. Robustheit als Prädiktorvariablen voraussagen lassen. Bevor wir jedoch diese Fragestellung mit Hilfe multipler Korrelationsanalysen bearbeiten, seien an dieser Stelle noch einige Bemerkungen zu den instrumentellen Eigenschaften der Prädiktor- und Kriteriumsvariablen gestattet.

Wenngleich Schulnoten eine Schätzskala von u n k l a r e r Metrik konstituieren, ist es nach TENT (1969, S. 75) "doch gerechtfertigt, Intervallskalenniveau in hinreichender Annäherung für Zeugnisnoten vorauszusetzen". Wie bereits in Kap. 2.7. auführlich erörtert, genügen Schulnoten nicht annähernd den Testgütekriterien: Die Objektivität der Noten ist mangelhaft; wie viele Untersuchungen belegen, kann es bei objektiv gleichen Leistungen zu erheblich voneinander abweichenden Beurteilungen kommen. Darüber hinaus sind Schulnoten besonders dann unbefriedigende Prädiktoren ihrer selbst, wenn zwischen den Schuljahren Lehrerwechsel liegen. Erteilt derselbe Lehrer die Noten für die verschiedenen Zeugnisse, so liegen nach RANK (1962) die Reliabilitätskoeffizienten von Jahr zu Jahr in der Größenordnung um .80, innerhalb eines Schuljahres um .90.

Hieraus wird die Tendenz erkennbar, einmal gefaßte Urteile nur noch wenig zu korrigieren, und es liegt nahe, zu vermuten, daß diese hohen Koeffizienten neben dem sog. Halo-Effekt auf implizite Persönlichkeitskonzepte des Lehrers zurückgehen. Diese Annahme wird gestützt durch die Tatsache, daß die in unserer Untersuchung als Prädiktoren betrachteten fünf Urteilsdimensionen nicht nur für die Gesamtstichprobe von N = 810 beurteilten Schülern hohe Reliabilitätsschätzungen aufweisen (S.54, Tab. 4b), sondern auch für die einzelnen Klassen eine beachtlich hohe innere Konsistenz ihrer Meßfunktion behalten (vergl. die Reliabilitätskoeffizienten in den Tabellen 1 bis 36 im Anhang). Die mangelnde Konstanz der Noten zwischen verschiedenen Schuljahren infolge Lehrerwechsels dürfte, was durch multiple Korrelationsanalysen überprüft werden soll, darauf zurückzuführen sein, daß verschiedene Lehrer die fünf Urteilsdimensionen im Rahmen des impliziten Persönlichkeitsmodells mit unterschiedlicher Gewichtung in ihrer Notengebung realisieren.

4.3.3.2. Darstellung globaler Zusammenhänge zwischen den Prädiktor- und Kriteriumsvariablen in einzelnen Klassen

Im folgenden ist nun zu zeigen, welche Beziehungen sich maximal zwischen den Prädiktor- und den Kriteriumsvariablen herstellen lassen. Zu diesem Zweck verwenden wir den multiplen Korrelationskoeffizienten R als Maß für die Enge des Zusammenhanges zwischen den fünf Urteilsdimensionen als Prädiktoren einerseits und der Schulnote in einem bestimmten Fach als Kriterium andererseits. Der multiple Korrelationskoeffizient gibt also die Korrelation zwischen einer Kriteriumsvariablen und einer gewichteten Summe von Prädiktorvariablen an und wird wie folgt berechnet:

$$R_{c.123\ldots k} = \sqrt{r_{c1}\beta_{c1} + r_{c2}\beta_{c2} + \ldots + r_{ck}\beta_{ck}}$$

In dieser Formel bedeuten:

$R_{c.123\ldots k}$ = Multipler Korrelationskoeffizient einer Batterie mit k Prädiktoren

$r_{c1}, r_{c2} \ldots r_{ck}$ = Korrelationskoeffizienten der einzelnen Prädiktoren mit dem Kriterium

$\beta_{c1}, \beta_{c2} \ldots \beta_{ck}$ = Standardpartial-Regressionskoeffizienten (Beta-Gewichte) der einzelnen Prädiktoren

Unsere multiplen Korrelationskoeffizienten wurden mit einem BMD/02R-Programm unter dem Namen stepwise regression errechnet, das, von der höchsten Korrelation eines Prädiktors mit dem Kriterium ausgehend, die weiteren Prädiktoren schrittweise so auswählt, daß nach jedem Schritt ein Maximum an zusätzlicher Varianz des Kriteriums aufgeklärt wird.

Tabelle 6 enthält für einzelne Klassen die multiplen Korrelationskoeffizienten zwischen den fünf Urteilsdimensionen (den Prädiktoren) und den letzten Zeugnisnoten in den Fächern Deutsch, Mathematik, Biologie, Physik, Englisch, Geschichte und Erdkunde (den Kriteriumsvariablen). Die Frage nach der statistischen Absicherung der mutiplen Korrelationskoeffizienten ist insofern schwierig, als die Koeffizienten auf sehr kleinen Stichproben basieren (Größenordnung zwischen N=10 und N=37). Die statistische Signifikanzprüfung erfolgte über $F = \frac{R^2(N-3)}{2(1-R^2)}$.

Tabelle 6

Multiple Korrelationskoeffizienten zwischen den fünf Beurteilungsdimensionen einerseits und Schulnoten in verschiedenen Fächern andererseits
(S = Sonderschule; H = Hauptschule; R = Realschule)

Lehrer Nr.	Geschlecht des Lehrers	Schulart	Schuljahr	Schüleranzahl	Deutsch Gesamt	Deutsch ♀	Deutsch ♂	Mathematik Gesamt	Mathematik ♀	Mathematik ♂	Biologie Gesamt	Biologie ♀	Biologie ♂	Physik Gesamt	Physik ♀	Physik ♂	Englisch Gesamt	Englisch ♀	Englisch ♂	Geschichte Gesamt	Geschichte ♀	Geschichte ♂	Erdkunde Gesamt	Erdkunde ♀	Erdkunde ♂
1	♂	S	7	12	.94*			.42																	
2	♀	S	8	14	.72*			.85*																	
3	♂	S	8	18	.83*			.86*																	
4	♂	S	7	14	-			-																	
5	♀	S	9	17	.87*			.72*																	
6	♀	S	8	10	.80*			.99*																	
7	♀	S	9	15	.62			.74*																	
8	♂	S	8	15	.82*			.67*																	
9	♂	S	8	14	-			-																	
10	♀	S	8	15	-			-																	
11	♀	S	7	24	.90*			.55*																	
12	♂	S	8	13	.89*			.77*																	
13	♀	S	9	16	.74*			.89*																	
14	♂	S	8	13	.69			.78*																	
15	♂	S	9	18	.46			.84*																	
16	♂	H	10	26	-			-			.63*						.72*								
17	♂	H	7	26							.76*	.86*	.84*				.70*	.77*	.82*						
18	♀	H	9	24	.86*	.80*	.85*																		
19	♀	H	8	29	.72*	.77*	.92*													.72*	.53	-			
20	♀	H	9	22	.41	.60	.74*													.64*	.61	.76*			
21	♀	H	7	36	.84*	.43	.92*				.54*	.44	.63												
22	♂	H	7	36	.65*	.71*	.53																.68*	.67*	.77*
23	♀	H	7	25	.60*	.79*	.59																		
24	♀	H	7	21	.75*	.85*	.90*																		
25	♂	H	7	23	.91*	.99*	.94*				.81*	.98*	.84*												
26	♂	H	9	22				.64*	.90*	.56															
27	♂	H	8	22	.76*	.95*	.89*	.90*	.83*	.97*															
28	♀	R	7	37	.57*	.68*	.62*										.50*	.50	.56						
29	♂	R	8	25				.77*	.77*	.96*				.43	.38	.75*									
30	♂	R	7	37				.68*	.70*	.72*				.64*	.88*	.70*									
31	♂	R	9	31													.67*	.74*	.71*	.49*	.67*	.72*			
32	♂	R	9	21				.82*						.79*											
33	♂	R	7	35													.66*								
34	♂	R	8	27																					
35	♂	R	8	30																					
36	♀	R	8	28							.77*	.69	.87*				.70*	.74*	.82*						
				$\bar{R}$	.76	.77	.84	.78	.79	.85	.70	.79	.80	.63	.77	.72	.65	.67	.74	.62	.62	.75	.68	.67	.77
Signifikanzniveau der Homogenitätsprüfung					.01	.01	.21	.01	n.s.	.01	n.s.	.01	n.s.	.01	.05	n.s.	.01	.01	.05	.01	.01	n.s.	.01	.01	.01

Von den 57 errechneten Koeffizienten erwiesen sich 51 als signifikant. Hieraus folgt, daß bedeutsame Zusammenhänge zwischen den fünf Urteilsdimensionen und den Schulnoten nachgewiesen werden konnten.

Für die Haupt- und Realschulklassen, die ein höheres N als die Sonderschulklassen aufweisen, wurden zudem die Koeffizienten getrennt nach Jungen und Mädchen errechnet. Von den dabei anfallenden Koeffizienten erwiesen sich 42 als siginifikant.

Zwecks besserer Übersichtlichkeit und Vergleichbarkeit wurden die Koeffizienten getrennt nach Fächern sowie nach Jungen und Mädchen über z' gemittelt, obgleich sich die Verteilungen der Koeffizienten in einigen Fällen (vergl. Tab. 6) als heterogen herausstellten. Die Homogenitätsprüfung erfolgte über

$$\chi^2 = \sum_{j=1}^{c} (N_j - 3) \cdot z'^2_j - \frac{\left[\sum_{j=1}^{c} (N_j - 3) \cdot z'_j\right]^2}{\sum_{j=1}^{c} (N_j - 3)}$$

Symbole:

$j = 1,2,\ldots,c$		Numerierung der Stichproben bzw. Korrelationskoeffizienten
c		Anzahl der Stichproben bzw. Korrelationskoeffizienten
N_j		Umfang der Stichprobe j
z'_j		geschätzter z-Wert für die Korrelation j nach McNEMAR (1962, S. 426)
df		Freiheitsgrade (degree of freedom)

Eine Inspektion der gemittelten Koeffizienten läßt vermuten, daß das Geschlecht als Moderatorvariable einen Einfluß auf die Höhe der multiplen Korrelation hat: Die getrennt nach Jungen und Mädchen errechneten Koeffizienten liegen in der Regel über denen, die sich aufgrund nicht gesplitterter Klassen ergeben, zudem zeigt sich ein Trend, daß die Koeffizienten bei Jungen höher ausfallen als bei Mädchen.

Wie die vorangegangenen multiplen Korrelationsanalysen gezeigt haben, lassen sich statistisch bedeutsame Zusammenhänge zwischen den fünf Urteilsdimensionen einerseits und den Schulnoten andererseits bei einzelnen Lehrern sichern. Dadurch ist Frage 3α(s. S.47) eindeutig beantwortet. Zudem sind die von uns gefundenen Validitätskoeffizienten in der Größenordnung von 0,7 bis 0,8 nach LIENERT als hoch bis sehr hoch zu bezeichnen. "Vom Blickpunkt der statistischen Vorhersage müssen Validitätskoeffizienten von $r_{tc} = 0,7$ verlangt werden, in der Praxis ist man jedoch schon mit Validitätskoeffizienten um 0,6 sehr zufrieden." (LIENERT 1967[2], S. 310).

4.3.3.3. Darstellung der schrittweisen Varianzaufklärung der Kriteriumsvariablen durch die Prädiktorvariablen in einzelnen Klassen

An dieser Stelle sei rückblickend betont, daß wir vor Beginn der Datenerhebung keine spezifizierten Hypothesen über den Zusammenhang der in der gegenwärtigen Untersuchung erfaßten Variablen aufgestellt hatten. Dennoch lautete unsere globale empirische Erwartung, daß überhaupt zwischen den 5 Urteilsdimensionen zur aktuellen Schülereinschätzung einerseits und den Zeugnisnoten andererseits bedeutsame Beziehungen bestehen. Die Existenz dieser allgemeinen Erwartung und das Fehlen spezifischer Hypothesen bestimmte

unser bisheriges Vorgehen: Zunächst schauten wir, ob tatsächlich Beziehungen zwischen den 5 Urteilsdimensionen sowie den Noten vorhanden sind (vgl. Frage 3a, S. 47). Angesichts der nachgewiesenen Beziehungen werden wir nun versuchen, durch entsprechende Spezifikationen und durch die Hinzunahme des Schülergeschlechts als Moderatorvariable die Art der Beziehung genauer zu erfassen. Frage 3b (siehe S. 47) stellt bereits eine Spezifikation zu Frage 3a dar. Als weitere Spezifikation läßt sich Frage 4a dahingehend formulieren, in welcher Reihenfolge und mit welchem Beitrag die fünf Urteilsdimensionen die Notenvarianz beim einzelnen Lehrer aufklären. Frage 4b zielt darauf ab, ob sich die fünf Urteilsdimensionen in ihrer Reihenfolge und in ihrem Beitrag zur Varianzaufklärung der Noten unter dem Aspekt der differentiellen Gültigkeit beim einzelnen Lehrer ändern; d.h. die schrittweise Varianzaufklärung der Noten durch die fünf Urteilsdimensionen wird bei den Real- und Hauptschullehrern für Jungen und Mädchen getrennt vorgenommen. Die Tabellen 7.1. bis 7.36 weisen die Ergebnisse der multiplen Korrelationsanalysen in den Klassen 1 bis 36 aus. Die kritischen F-Werte zur Prüfung signifikanter Erhöhungen der multiplen Korrelationskoeffizienten sind im Programmausdruck mit enthalten. Signifikante Erhöhungen sind von uns mit einem Stern versehen.

Bei Inspektion aller 112 Koeffizienten fällt auf, daß zur schrittweisen Aufklärung der Notenvarianzen die Dimension III (Begabung) 54mal und die Dimension I (Arbeitsverhalten) 41mal an erster Stelle stehen. Aufgrund zu geringer Stichprobenumfänge lassen sich für die Koeffizienten viele Zuwächse nicht sichern. Dies trifft insbesondere für die Sonderschulklassen und die Teilstichproben (Jungen und Mädchen) zu. An den multiplen Korrelationskoeffizienten ist nicht zu erkennen, welche der fünf Prädiktoren positiv und welche negativ mit dem Kriterium korrelieren. Deshalb ist es notwendig, die

Interkorrelationsmatrizen zur Interpretation der multiplen Korrelationskoeffizienten aus den Rechnungsprotokollen mit heranzuziehen. Die Interkorrelationsmatrizen sind aus Platzgründen nicht mit aufgeführt.

Tabelle 7.1 bis 7.36

Ergebnisse der multiplen Korrelationsanalysen in den einzelnen Klassen

Legende

P: Prädiktorvariablen (I= Disziplinertheit vs. Undisziplinertheit;
II= Soziale Aufgeschlossenheit vs.Verschlossenh
III= Geistige Regsamkeit;
IV= Normenkonformität;
V= Empfindsamkeit vs. Robustheit)

R: Multiple Korrelationskoeffizienten

R^2: Quadrierte Korrelationskoeffizienten

F: Prüfgröße zur Ermittlung signifikanter Erhöhungen, die mit einem Stern (*) versehen sind

Lehrer I

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F
III	.85	.72	25.29 *
I	.92	.85	7.88 *
II	.94	.88	1.68
IV	.94	.89	0.84
V	.94	.89	0.11

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F
V	.36	.13	1.52
III	.40	.16	0.28
IV	.40	.16	0.22
I	.42	.17	0.11
II	.42	.17	0.00

Lehrer II (Überschrift wie vor)

P	R	R^2	F
III	.51	.26	4.20
V	.67	.46	3.96
IV	.68	.47	0.21
II	.72	.51	0.85
I	.72	.51	0.03

P	R	R^2	F
III	.57	.32	5.75 *
IV	.79	.62	8.68 *
I	.81	.66	1.15
V	.84	.71	1.49
II	.85	.73	0.55

Lehrer III

Kriteriumvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.68	.46	13.87	
III	.73	.54	2.39	
V	.82	.67	5.75	
II	.82	.67	0.06	
IV	.83	.69	0.47	

Kriteriumvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.73	.53	18.12	*
V	.84	.70	8.28	*
II	.89	.73	1.52	
I	.86	.74	0.51	
IV	.86	.74	0.02	

Lehrer V

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.57	.33	7.33	*
IV	.83	.69	16.24	*
II	.87	.75	3.20	
III	.87	.76	0.51	
V	.87	.76	0.03	

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.29	.29	6.12	*
I	.67	.44	3.89	*
II	.72	.52	2.11	
V	.72	.52	0.00	
IV	.72	.52	0.00	

Lehrer VI

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.67	.45	6.51	*
IV	.76	.57	2.06	
III	.80	.63	.98	
II	.80	.64	.11	
V	.80	.64	.01	

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.89	.80	31.92	*
V	.92	.85	2.19	
II	.93	.86	.46	
I	.99	.97	20.60	
IV	.99	.97	.00	

Lehrer VII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
IV	.39	.15	2.34	
V	.48	.23	1.25	
II	.61	.37	2.37	
III	.61	.37	.08	
I	.62	.38	.15	

Kriteriumsvariable:Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.50	.25	4.24	
V	.67	.44	4.27	*
IV	.69	.48	.81	
I	.73	.54	1.25	
II	.74	.55	.11	

Lehrer VIII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.72	.52	12.88	*
V	.81	.66	4.62	*
IV	.81	.66	.06	
III	.82	.67	.11	
II	.82	.67	.03	

Kriteriumsvariable:Mathematinote

Gesamtstichprobe

P	R	R^2	F	
III	.45	.20	3.08	
IV	.66	.44	4.58	*
I	.67	.45	.22	
II	.67	.45	.02	
V	.67	.45	.00	

Lehrer XI

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.83	.69	48.96	*
III	.88	.77	6.85	*
V	.90	.80	3.61	*
II	.90	.80	.03	
IV	.90	.80	.00	

Kriteriumsvariable:Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
I	.45	.20	5.51	*
II	.51	.26	1.70	
IV	.54	.29	.97	
III	.55	.30	.18	
V	.55	.30	.00	

Lehrer XII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.70	.48	10.29	*
I	.73	.53	.92	
V	.80	.63	2.61	
I	.86	.75	3.50	
V	.89	.79	1.49	

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.58	.33	5.44	*
I	.74	.55	4.84	*
V	.75	.56	.31	
II	.77	.56	.31	
IV	.77	.59	.00	

Lehrer XIII

Kriteriumsvariable: Deutschnote

Ges-Stichprobe

P	R	R^2	F	
I	.68	.47	12.31	*
I	.71	.51	1.00	
I	.73	.53	.63	
V	.74	.54	.28	
V	.74	.55	.10	

Kriteriumsvariable: Mathematiknote

Ges.-Stichprobe

P	R	R^2	F	
I	.64	.41	9.73	*
III	.67	.45	.90	
IV	.76	.58	3.65	*
V	.87	.75	7.70	*
II	.89	.79	2.09	

Lehrer XIV

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
V	.31	.09	1.15	
V	.45	.20	1.37	
I	.63	.39	2.78	
I	.69	.47	1.20	
I	.69	.48	.14	

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
I	.61	.37	6.47	*
V	.64	.41	.69	
IV	.72	.52	1.96	
II	.76	.58	1.14	
III	.78	.62	.71	

Lehrer XV

Kriteriumsvariable: Deutschnote
Gesamtstichprobe

P	R	R^2	F
I	.33	.11	1.90
IV	.42	.17	1.22
III	.44	.19	.32
II	.45	.20	.14
V	.46	.22	.23

Kriteriumsvariable: Mathematiknote
Gesamtstichprobe

P	R	R^2	F
III	.63	.39	10.26 *
I	.78	.60	7.98 *
V	.81	.66	2.49
II	.83	.69	1.32
IV	.84	.70	.29

Lehrer XVI

Kriteriumsvariable: Englischnote
Gesamtstichprobe

P	R	R^2	F
III	.65	.42	17.23
I	.66	.43	.59
II	.67	.44	.40
V	.71	.51	2.80
IV	.72	.51	.27

Kriteriumsvariable: Biologienote
Gesamtstichprobe

P	R	R^2	F
I	.57	.32	11.26 *
V	.62	.38	2.37
IV	.63	.39	.28
III	.63	.39	.03
II	.63	.39	.03

Lehrer XVII

Kriteriumsvariable: Englischnote
Gesamtstichprobe

P	R	R^2	F
I	.66	.44	18.17 *
III	.68	.46	.63
II	.70	.49	1.43
V	.70	.49	.09
IV	.70	.50	.09

Kriteriumsvariable: Biologienote
Gesamtstichprobe

P	R	R^2	F
I	.71	.51	23.50 *
IV	.73	.53	1.07
V	.74	.55	.82
III	.75	.56	.84
II	.76	.58	.61

Mädchen

P	R	R^2	F	
V	.60	.36	6.29	*
III	.73	.53	3.67	
IV	.74	.55	.38	
I	.75	.56	.20	
II	.77	.60	.61	

Jungen

P	R	R^2	F	
I	.73	.54	11.64	*
II	.75	.56	.47	
V	.77	.60	.70	
III	.81	.66	1.26	
IV	.82	.68	.42	

Mädchen

P	R	R^2	F	
I	.84	.70	25.95	*
IV	.86	.74	1.29	
II	.86	.74	.09	
III	.86	.74	.08	
V	.86	.74	.04	

Jungen

P	R	R^2	F	
IV	.63	.39	6.45	*
V	.79	.62	5.32	*
I	.83	.70	2.08	
II	.84	.71	.26	
III	.84	.70	.00	

Lehrer XVIII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
III	.78	.61	34.40	*
I	.84	.71	7.37	*
II	.86	.73	1.69	
V	.86	.73	.08	
IV	.86	.74	.01	

Mädchen

P	R	R^2	F	
V	.54	.29	5.69	*
III	.63	.39	2.23	
IV	.74	.55	4.13	
I	.78	.61	1.82	
II	.80	.64	.77	

Jungen

I	.48	.23	1.86	
III	.56	.31	.54	
IV	.74	.54	2.02	
II	.81	.66	1.05	
V	.85	.72	.41	

Kriteriumsvariable: Biologienote

Gesamtstichprobe

P	R	R^2	F

Lehrer XIX

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
III	.52	.28	10.27	*
IV	.68	.46	9.13	*
V	.69	.48	.71	
II	.70	.49	.73	
I	.72	.52	1.02	

Mädchen

P	R	R^2	F
III	.70	.49	10.51
IV	.73	.54	1.07
I	.74	.55	.29
V	.76	.58	.43
II	.77	.60	.36

Jungen

P	R	R^2	F
III	.87	.76	44.42
I	.90	.82	3.63
II	.91	.83	1.52
V	.92	.84	.73
IV	.92	.85	.48

Kriteriumsvariable: Geschichtsnote

Gesamtstichprobe

P	R	R^2	F	
I	.45	.20	6.93	*
II	.56	.31	4.08	*
V	.67	.45	6.30	*
III	.70	.49	1.78	
IV	.72	.52	1.58	

Mädchen

P	R	R^2	F
I	.47	.22	3.15
II	.49	.24	.23
III	.50	.25	.14
V	.53	.28	.35
IV	.54	.29	.06

Jungen

P	R	R^2	F
I	.00	.00	.00
II	.00	.00	.00
III	.00	.00	.00
IV	.00	.00	.00
V	.00	.00	.00

Lehrer XX

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F
III	.35	.12	2.85
IV	.38	.14	.43
I	.39	.15	.15
II	.40	.16	.14
V	.41	.17	.23

Kriteriumsvariable: Geschichtsnote

Gesamtstichprobe

P	R	R^2	F	
IV	.46	.21	5.32	*
I	.51	.26	1.32	
V	.54	.29	.64	
III	.61	.37	2.24	
II	.64	.41	1.09	

Mädchen

P	R	R^2	F
I	.36	.13	1.53
V	.47	.22	1.03
II	.58	.33	1.34
IV	.60	.36	.27
III	.60	.36	.00

Mädchen

P	R	R^2	F
I	.39	.15	1.82
II	.43	.18	.34
III	.57	.32	1.59
V	.61	.38	.62
IV	.61	.38	.00

Jungen

P	R	R^2	F
I	.50	.25	2.67
IV	.67	.45	2.48
III	.74	.55	1.31
V	.74	.55	.03
II	.74	.55	.00

Jungen

P	R	R^2	F
IV	.66	.43	6.05 *
I	.75	.57	2.17
III	.76	.57	.09
V	.76	.58	.12
II	.76	.58	.02

Lehrer XXI

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F
I	.80	.64	60.21 *
IV	.82	.68	4.24 *
II	.84	.70	1.92
V	.84	.70	.15
III	.84	.70	.06

Kriteriumsvariable: Biologienote

Gesamtstichprobe

P	R	R^2	F
I	.43	.18	7.53 *
IV	.50	.25	2.86
III	.51	.26	.68
V	.54	.29	1.37
II	.54	.29	.00

Mädchen

P	R	R^2	F
I	.36	.13	2.45
II	.41	.17	.59
V	.42	.18	.19
IV	.42	.18	.05
III	.43	.18	.06

Mädchen

P	R	R^2	F
III	.17	.03	.50
I	.23	.05	.33
IV	.41	.17	1.98
V	.43	.18	.19
II	.44	.20	.25

(Fortsetzung Lehrer XXI)

Jungen

P	R	R^2	F	
I	.88	.78	57.44	*
IV	.92	.85	6.53	*
III	.92	.85	.06	
V	.92	.85	.01	
II	.92	.85	.00	

Jungen

P	R	R^2	F	
I	.55	.30	6.85	*
II	.61	.37	1.70	
III	.63	.40	.63	
IV	.63	.40	.06	
V	.63	.40	.03	

Lehrer XXII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

	R	R^2	F	
I	.63	.40	23.09	*
I	.64	.41	.15	
V	.64	.41	.13	
V	.65	.42	.35	
I	.65	.42	.06	

Mädchen

	R	R^2	F	
I	.68	.46	17.96	
I	.69	.47	.33	
V	.69	.48	.42	
V	.70	.49	.48	
I	.71	.50	.17	

Jungen

	R	R^2	F	
I	.51	.26	3.87	
I	.52	.27	.09	
V	.52	.27	.09	
I	.53	.28	.01	
V	.53	.28	.00	

Kriteriumsvariable: Erdkundenote

Gesamtstichprobe

	R	R^2	F	
I	.58	.33	17.03	*
II	.64	.41	4.14	*
IV	.67	.45	2.64	
III	.68	.46	.59	
V	.68	.46	.05	

Mädchen

	R	R^2	F	
I	.61	.38	12.61	
V	.64	.40	.95	
IV	.64	.41	.23	
II	.67	.45	1.19	
III	.67	.45	.09	

Jungen

	R	R^2	F	
III	.52	.27	4.15	
V	.72	.52	5.03	*
II	.75	.56	.77	
IV	.76	.58	.57	
I	.77	.59	.04	

Lehrer XXIII

Kriteriumsvariable : Deutschnote

Gesamtstichprobe

	R	R^2	F	
II	.45	.21	5.74	*
II	.55	.31	3.02	
I	.58	.34	1.06	
V	.60	.36	.51	

Mädchen

	R	R^2	F	
I	.48	.23	2.64	
II	.64	.41	2.42	
IV	.72	.53	1.76	
II	.79	.62	1.43	
V	.79	.62	.04	

Jungen

	R	R^2	F	
I	.17	.03	.31	
III	.43	.18	1.88	
IV	.58	.33	2.01	
II	.59	.35	.18	
V	.59	.35	.04	

Lehrer XXIV

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
III	.57	.33	9.22	*
V	.63	.40	2.05	
IV	.66	.44	1.31	
I	.71	.51	2.21	
II	.75	.56	1.98	

Mädchen

P	R	R^2	F	
III	.70	.50	6.90	*
II	.73	.53	.43	
V	.83	.69	2.51	
I	.85	.72	.50	
IV	.85	.72	.00	

Jungen

P	R	R^2	F	
III	.43	.19	2.29	
V	.53	.29	1.25	
IV	.70	.49	3.25	
I	.87	.76	7.85	*
II	.90	.80	1.26	

Lehrer XXII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

R	R^2	F	
.63	.40	23.09	*
.64	.41	.15	
.64	.41	.13	
.65	.42	.35	
.65	.42	.06	

Mädchen

R	R^2	F
.68	.46	17.96
.69	.47	.33
.69	.48	.42
.70	.49	.48
.71	.50	.17

Jungen

R	R^2	F
.51	.26	3.87
.52	.27	.09
.52	.27	.09
.53	.28	.01
.53	.28	.00

Kriteriumsvariable: Erdkundenote

Gesamtstichprobe

P	R	R^2	F	
I	.58	.33	17.03	*
II	.64	.41	4.14	*
IV	.67	.45	2.64	
III	.68	.46	.59	
V	.68	.46	.05	

Mädchen

P	R	R^2	F
I	.61	.38	12.61
V	.64	.40	.95
IV	.64	.41	.23
II	.67	.45	1.19
III	.67	.45	.09

Jungen

P	R	R^2	F	
III	.52	.27	4.15	
V	.72	.52	5.03	*
II	.75	.56	.77	
IV	.76	.58	.57	
I	.77	.59	.04	

Lehrer XXIII

Kriteriumsvariable : Deutschnote

Gesamtstichprobe

R	R^2	F	
.45	.21	5.74	*
.55	.31	3.02	
.58	.34	1.06	
.60	.36	.51	

Mädchen

R	R^2	F
.48	.23	2.64
.64	.41	2.42
.72	.53	1.76
.79	.62	1.43
.79	.62	.04

Jungen

P	R	R^2	F
I	.17	.03	.31
III	.43	.18	1.88
IV	.58	.33	2.01
II	.59	.35	.18
V	.59	.35	.04

Lehrer XXIV

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
III	.57	.33	9.22	*
V	.63	.40	2.05	
IV	.66	.44	1.31	
I	.71	.51	2.21	
II	.75	.56	1.98	

Mädchen

P	R	R^2	F	
III	.70	.50	6.90	*
II	.73	.53	.43	
V	.83	.69	2.51	
I	.85	.72	.50	
IV	.85	.72	.00	

Jungen

P	R	R^2	F	
III	.43	.19	2.29	
V	.53	.29	1.25	
IV	.70	.49	3.25	
I	.87	.76	7.85	*
II	.90	.80	1.26	

Lehrer XXV

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
I	.81	.66	40.33	*
I	.87	.75	7.87	*
V	.90	.81	6.08	*
V	.91	.82	.80	
I	.91	.82	.01	

Mädchen

P	R	R^2	F	
I	.90	.81	25.31	*
I	.92	.84	.91	
V	.95	.91	3.18	
V	.99	.98	12.59	*
I	.99	.99	1.90	

Jungen

P	R	R^2	F	
I	.80	.63	22.57	*
I	.88	.78	8.13	*
I	.92	.85	4.60	*
V	.93	.86	1.08	
V	.94	.88	1.40	

Kriteriumsvariable: Biologienote

Gesamtstichprobe

P	R	R^2	F	
I	.72	.52	22.36	*
III	.79	.62	5.27	*
V	.79	.62	.19	
II	.80	.65	1.32	
IV	.81	.65	.23	

Mädchen

P	R	R^2	F	
IV	.96	.92	66.69	*
I	.97	.95	3.09	
V	.97	.95	.11	
III	.98	.95	.01	
II	.98	.95	.01	

Jungen

P	R	R^2	F	
III	.77	.60	19.52	*
I	.80	.63	1.06	
II	.81	.66	.72	
V	.84	.71	1.71	
IV	.84	.71	.09	

Lehrer XVI

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
I	.55	.30	8.50	*
V	.61	.37	2.04	
I	.63	.39	.76	
II	.64	.41	.52	
IV	.64	.41	.04	

Mädchen

P	R	R^2	F	
I	.83	.69	18.14	*
II	.85	.72	.77	
III	.87	.76	1.03	
V	.90	.81	1.29	
IV	.90	.82	.05	

Jungen

P	R	R^2	F
III	.43	.18	2.22
I	.54	.29	1.35
IV	.55	.30	.15
V	.56	.31	.00

Lehrer XXVII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
V	.53	.28	7.66	*
III	.68	.46	6.27	*
II	.75	.56	4.47	*
I	.76	.57	.26	
IV	.76	.57	.00	

Mädchen

P	R	R^2	F	
IV	.73	.54	10.39	*
II	.83	.70	4.26	*
III	.92	.85	7.30	*
I	.94	.89	1.74	
V	.95	.90	.50	

Jungen

P	R	R^2	F	
V	.60	.36	5.07	*
III	.73	.53	2.90	
I	.80	.64	2.03	
II	.85	.72	1.68	
IV	.89	.78	1.59	

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.09	.77	68.56	*
IV	.89	.80	2.50	
I	.90	.81	.78	
II	.90	.81	.54	
V	.90	.81	.03	

Mädchen

P	R	R^2	F	
III	.81	.66	17.13	*
V	.82	.67	.23	
I	.82	.67	.21	
IV	.83	.69	.21	
II	.83	.69	.03	

Jungen

P	R	R^2	F	
III	.94	.88	65.64	*
IV	.97	.93	6.66	*
II	.97	.94	.17	
I	.97	.94	.02	
V	.97	.94	.01	

Lehrer XXVIII

Kriteriumsvariable: Deutschnote

Gesamtstichprobe

P	R	R^2	F	
III	.53	.29	13.96	*
IV	.56	.32	1.49	
II	.57	.33	.50	
I	.57	.33	.12	
V	.57	.33	.01	

Kriteriumsvariable: Englischnote

Gesamtstichprobe

P	R	R^2	F	
I	.40	.16	6.49	*
III	.45	.20	1.94	
IV	.48	.23	1.41	
V	.50	.25	.57	
II	.50	.25	.01	

Mädchen

P	R	R^2	F	
II	.61	.37	12.59	*
IV	.62	.39	.45	
II	.65	.42	1.05	
V	.67	.46	1.15	
I	.68	.46	.28	

Mädchen

P	R	R^2	F	
III	.44	.20	5.15	*
II	.49	.24	1.10	
IV	.50	.25	.22	
V	.50	.25	.01	
I	.50	.25	.00	

Jungen

P	R	R^2	F	
II	.60	.36	6.65	*
V	.62	.38	.50	
II	.62	.39	.03	
I	.62	.39	.01	
IV	.62	.39	.00	

Jungen

P	R	R^2	F
IV	.40	.16	2.33
I	.54	.29	2.01
II	.54	.30	.06
III	.56	.31	.16
V	.56	.31	.00

Lehrer XXIX

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.57	.33	11.17	*
II	.68	.47	5.72	*
V	.76	.58	5.65	*
I	.77	.59	.35	
IV	.77	.59	.00	

Kriteriumsvariable: Physiknote

Gesamtstichprobe

P	R	R^2	F	
III	.41	.16	4.52	*
II	.41	.17	.12	
V	.42	.18	.30	
IV	.43	.19	.15	
I	.43	.19	.00	

Mädchen

P	R	R^2	F	
III	.58	.34	6.17	*
II	.68	.47	2.65	
I	.76	.58	2.72	
V	.77	.59	.28	
IV	.77	.59	.00	

Mädchen

P	R	R^2	F
IV	.31	.10	1.27
I	.33	.11	.13
V	.34	.12	.12
II	.38	.15	.31
III	.38	.15	.01

Jungen

P	R	R^2	F	
III	.64	.41	6.30	*
V	.79	.62	4.38	*
I	.87	.75	3.84	
II	.96	.93	14.11	*
IV	.96	.93	.00	

Jungen

P	R	R^2	F
III	.56	.32	4.17
V	.59	.35	.42
I	.69	.47	1.56
II	.75	.56	1.17
IV	.75	.56	.07

Lehrer XXX

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.53	.28	13.91	*
V	.64	.42	7.62	*
II	.67	.46	2.43	
I	.68	.46	.14	
IV	.68	.46	.24	

Kriteriumsvariable: Physiknote

Gesamtstichprobe

P	R	R^2	F	
III	.60	.35	19.22	*
II	.63	.40	2.40	
I	.64	.42	1.06	
V	.64	.42	.00	
IV	.64	.42	.00	

Mädchen

P	R	R^2	F	
III	.60	.36	8.31	*
II	.64	.42	1.42	
V	.67	.45	.80	
IV	.69	.48	.65	
I	.70	.49	.30	

Mädchen

P	R	R^2	F	
II	.60	.36	8.44	*
I	.85	.72	18.06	*
III	.87	.76	1.86	
V	.88	.77	.71	
IV	.88	.77	.05	

Jungen

P	R	R^2	F	
I	.55	.30	7.85	*
V	.69	.48	5.74	*
IV	.71	.50	.78	
III	.72	.51	.25	
II	.72	.51	.00	

Jungen

P	R	R^2	F	
III	.64	.41	12.61	*
IV	.69	.47	2.03	
I	.69	.48	.18	
II	.69	.48	.02	
V	.70	.48	.05	

Lehrer XXXI

Kriteriumsvariable: Englischnote

Gesamtstichprobe

P	R	R^2	F	
III	.58	.34	14.91	*
I	.66	.44	4.76	*
II	.67	.44	.38	
IV	.67	.45	.19	
V	.67	.45	.34	

Mädchen

P	R	R^2	F	
III	.58	.34	9.85	*
IV	.66	.43	2.92	
II	.68	.46	.81	
I	.73	.53	2.44	
V	.74	.55	.52	

Jungen

P	R	R^2	F
I	.62	.38	4.99
V	.69	.47	1.13
IV	.70	.48	.16
II	.71	.50	.15
III	.71	.50	.03

Kriteriumsvariable: Geschichtsnote

Gesamtstichprobe

P	R	R^2	F	
III	.33	.11	3.53	
V	.47	.22	3.95	*
IV	.49	.24	.65	
I	.49	.24	.07	
II	.49	.24	.00	

Mädchen

P	R	R^2	F	
I	.44	.20	4.61	*
V	.53	.28	2.04	
III	.60	.36	2.27	
IV	.61	.37	.17	
II	.61	.38	.20	

Jungen

P	R	R^2	F
III	.57	.33	3.89
I	.68	.46	1.66
II	.69	.47	.20
IV	.71	.50	.26
V	.72	.53	.22

Lehrer XXXII

Kriteriumsvariable: Mathematiknote

Gesamtstichprobe

P	R	R^2	F	
III	.73	.53	21.48	*
I	.79	.63	4.53	*
IV	.82	.66	2.00	
V	.82	.68	.61	
II	.82	.68	.08	

Kriteriumsvariable: Physiknote

Gesamtstichprobe

P	R	R^2	F	
III	.74	.54	22.52	*
V	.78	.61	2.96	
I	.79	.62	.65	
II	.79	.62	.01	
IV	.79	.62	.00	

Lehrer XXXVI

Kriteriumsvariable: Englischnote

Gesamtstichprobe

P	R	R^2	F	
I	.63	.39	16.67	*
II	.66	.44	2.23	
III	.69	.47	1.55	
IV	.69	.48	.24	
V	.70	.48	.16	

Mädchen

P	R	R^2	F	
V	.55	.30	4.75	
IV	.72	.52	4.49	*
II	.73	.54	.42	
I	.74	.54	.02	

Jungen

P	R	R^2	F	
I	.77	.60	19.24	*
III	.78	.61	.40	
II	.79	.62	.27	
V	.81	.65	1.00	
IV	.82	.67	.37	

Kriteriumsvariable: Biologienote

Gesamtstichprobe

P	R	R^2	F	
III	.70	.49	25.03	*
IV	.74	.54	2.95	
V	.76	.58	2.19	
I	.77	.59	.39	
II	.77	.59	.09	

Mädchen

P	R	R^2	F	
I	.64	.41	7.69	*
II	.66	.44	.52	
III	.68	.46	.31	
IV	.69	.48	.13	

Jungen

P	R	R^2	F	
III	.85	.72	34.07	*
V	.87	.76	1.49	
I	.87	.76	.29	
II	.87	.76	.13	
IV	.87	.76	.00	

Die Interpretation der Ergebnisse der Korrelationsanalysen soll wiederum nicht für alle Klassen erfolgen. Um die Interpretation der Ergebnisse bei einzelnen Lehrern zu demonstrieren, wählen wir Lehrer 5 (Nr. 3, 17, 18, 26, 29) aus.

Lehrer III

Bei Lehrer III handelt es sich um einen Sonderschullehrer, der in einer Erziehungsschwierigenklasse (N = 18) unterrichtet. Bei ihm wurden die Noten der Fächer Deutsch und Mathematik erhoben. Beide Notenvarianzen werden durch die fünf HOFER-Faktoren zur aktuellen Schülerbeurteilung sehr hoch aufgeklärt. (Deutsch : R = .83; Mathematik: R = .86). Eine gute Deutschnote bekommen diejenigen Schüler und Schülerinnen, die nach Ansicht des Lehrers ein gutes Arbeitsverhalten zeigen und robust und unkompliziert sind. In Mathematik erhalten diejenigen Schüler und Schülerinnen die besten Noten, die der Lehrer als begabt und ebenfalls robust und unkompliziert einschätzt. Auf eine getrennte Analyse nach Jungen und Mädchen mußten wir bei einer Klassenstärke von N = 18 verzichten.

Lehrer XVII

Bei Lehrer XVII handelt es sich um einen Hauptschullehrer, der ein siebtes Schuljahr unterrichtet. Bei ihm wurden die Noten der Fächer Englisch und Biologie erhoben. Beide Notenvarianzen werden durch die fünf HOFER-Faktoren hoch aufgeklärt. (Englisch: R = .70; Biologie: R = .76) Für die Gesamtstichprobe (N = 25) gilt, daß in beiden Fächern diejenigen Kinder die besten Noten erhalten, deren Arbeitsverhalten vom Lehrer hoch eingestuft wird. Erhebliche Unterschiede zeigen sich jedoch bei Mädchen und Jungen. Eine gute Englischnote erhalten diejenigen Mädchen, die der Lehrer als sensibel und empfindsam ansieht und diejenigen Jungen, deren Arbeitsverhalten er positiv beurteilt. Gute Biologienoten vergibt der Lehrer an Schülerinnen mit "gutem Arbeitsverhalten" und an Schüler, die sich normenkonform (im Sinne von höflich und folgsam) "verhalten" und die er als robust und unkompliziert einschätzt.

Lehrer XVIII

Bei Lehrer XVIII handelt es sich um eine Hauptschullehrerin, die Deutsch in einem neunten Schuljahr (N = 24) unterrichtet. Die Deutschnote wird bezogen auf die Gesamtstichprobe, sehr hoch aufgeklärt (R = .86). "Begabte" und "fleißige" Schüler und Schülerinnen erhalten gute Noten in Deutsch. Gute Deutschnoten korrelieren bei den Mädchen mit "Sensibilität". Zur Varianzaufklärung der Deutschnote bei den Jungen tragen die fünf HOFER-Faktoren zu etwa gleichen Teilen bei. Die einzelnen Zuwächse heben sich zwar nicht statistisch bedeutsam voneinander ab, insgesamt erhalten wir jedoch einen sehr hohen multiplen Koeffizienten von R = .85. Zudem wird bei den Jungen die Tendenz erkennbar, daß gute Deutschnoten nicht wie bei den Mädchen mit "Sensibilität" und Empfindsamkeit", sondern im Gegensatz dazu mit "Robustheit" zusammenhängen.

Lehrer XXVI

Bei Lehrer XXVI handelt es sich um einen Hauptschullehrer, der in einer neunten Kla
(N = 22) Mathematik unterrichtet. Die Mathematiknote wird, bezogen auf die Gesamt-
stichprobe, durchschnittlich hoch aufgeklärt (R = .64). Schüler und Schülerinnen mi
"hoher Begabung" erhalten gute Mathematiknoten. Bei den Mädchen sind gute Mathematik-
noten von "gutem Arbeitsverhalten" abhängig. Bei den Jungen läßt sich eine statistisc
bedeutsame Aufklärung der Varianz der Mathematiknote nicht nachweisen.

Lehrer XXIX

Bei Lehrer XXIX handelt es sich um einen Realschullehrer, der Mathematik und Physik
in einem achten Schuljahr (N = 25) unterrichtet. Die beiden Notenvarianzen werden
unterschiedlich hoch aufgeklärt (Mathematik: R = .77; Physik: R = .43). Für di
Gesamtstichprobe gilt, daß diejenigen Schüler und Schülerinnen gute Mathematiknoten
aufweisen, die der Lehrer für begabt, zurückhaltend und sensibel hält. Gute Physiknot
gehen nach Ansicht des Lehrers mit Begabung einher. Unterschiedliche Zusammenhänge z
den Urteilsdimensionen einerseits und der Mathematik- und Physiknote andererseits er
sich für Jungen und Mädchen. Bei den Jungen wird im Gegensatz zu den Mädchen die Mathe
note fast völlig und die Physiknote sehr hoch durch die HOFER- Faktoren aufgeklärt . B
den Mädchen läßt sich der multiple Koeffzient von .38 zwischen den HOFER-Faktoren
der Physiknote statistisch nicht sichern. Mädchen mit guten Mathematiknoten werden vom
Lehrer als begabt, Jungen als begabt und sensibel eingestuft. Die Dimension Zurückhalt
Geltungsstreben stellt zur Varianzaufklärung der Mathematiknote bei den Jungen eine so
nannte Suppressor-Variable (vergl. LIENERT 1967[2], S. 393) dar, die selbst mit der Mat
matiknote nicht korreliert, aber die multiple Korrelation insgesamt statistisch bedeuts
erhöht.

5. Zusammenfassung der Ergebnisse

Eine R-Analyse mit anschließender Varimax-Rotation der 25 polaren Eigenschaftspaare (vergl. Anhang S.123) zur aktuellen Schülerbeurteilung, der eine Stichprobe von 810 beurteilten Schülern und Schülerinnen zugrundelag, erbrachte (wie bei HOFER 1969) eine fünffaktorielle Aufgliederung als optimal interpretierbare Lösung. Die Faktoren bzw. Urteilsdimensionen wurden wir folgt benannt:

I. D i s z i p l i n i e r t h e i t vs. U n - d i s z i p l i n i e r t h e i t mit den Eigenschaften interessiert/uninteressiert, fleißig/faul, ehrgeizig/gleichgültig, unkonzentriert/konzentriert, unordentlich/ordentlich, pflichtbewußt/pflichtvergessen, aufmerksam/abgelenkt und zuverlässig/unzuverlässig.

II. Z u r ü c k h a l t u n g vs. G e l t u n g s - s t r e b e n mit den Eigenschaften verschlossen/offen, selbstsicher/selbstunsicher, gesellig/ungesellig, geltungsbedürftig/bescheiden, schüchtern/dreist, unruhig/ruhig und führend/zurückhaltend.

III. G e i s t i g e R e g s a m k e i t mit den Eigenschaften unbegabt/begabt, einfallsreich/einfallslos, intelligent/dumm und aufgeweckt/träge.

IV. N o r m e n k o n f o r m i t ä t mit den Eigenschaften höflich/frech, folgsam/unfolgsam, ausgeglichen/launisch und sympathisch/unsympathisch.

V. E m p f i n d s a m k e i t vs. R o b u s t - h e i t mit den Eigenschaften sensibel/robust und kompliziert/unkompliziert.

Bis auf Faktor IV, den wir als "Normenkonformität" interpretierten, konnte die Benennung der weiteren Faktoren nach HOFER (1969, S. 60) erfolgen.

Indem wir die Schülerprofile bezüglich der fünf Urteilsdimensionen in einzelnen Klassen mit der Hierarchical Grouping Analysis nach WARD (1963), die Unterschiede hinsichtlich der Lage (evaluation) zwischen Profilen und Varianzunterschiede (scatter) zwischen Profilen berücksichtigt, clusteranalysierten, erhielten wir von Lehrer zu Lehrer drei Schülergruppen mit sehr unterschiedlichen mittleren Ausprägungsgraden auf unterschiedlich vielen der fünf Urteilsdimensionen. Kriterien zur interpretativen Beschreibung der einzelnen Profilabbildungen wurden erarbeitet und einige ausgewählte Profilabbildungen interpretiert. Für alle Lehrer konnten statistisch bedeutsame Zusammenhänge zwischen den fünf Urteilsdimensionen einerseits und Schulnoten in verschiedenen Fächern andererseits nachgewiesen werden. Die gefundenen Validitätskoeffizienten in der Größenordnung von .7 bis .8 können als hoch bis sehr hoch bezeichnet werden. Eine getrennte Betrachtung der Zusammenhänge bei Jungen und Mädchen läßt vermuten, daß das Geschlecht als Moderatorvariable einen Einfluß auf die Höhe der multiplen Korrelation hat. Die für Jungen und Mädchen errechneten Koeffizienten liegen in der Regel über denen, die sich aufgrund nicht gesplitterter Klassen ergeben. Zudem zeigt sich ein Trend, daß die Koeffizienten bei Jungen höher ausfallen als bei Mädchen, daß bei Jungen also insgesamt mehr Notenvarianz durch die HOFER-Faktoren aufgeklärt wird als bei Mädchen.

Bei Inspektion aller 112 Koeffizienten fällt auf, daß zur schrittweisen Aufklärung der Notenvarianzen die Dimension Begabung 54 mal und die Dimension Arbeitsverhalten 41 mal an erster Stelle stehen. Aufgrund zu geringer Stichprobenumfänge lassen sich für die Koeffizienten viele Zuwächse statistisch nicht sichern. Dies trifft insbesonde-

re für die Koeffizienten zu, die aufgrund der Sonderschulklassen und der nach Jungen und Mädchen aufgeteilten Haupt- und Realschulklassen errechnet wurden.

Die Tabellen 7.1 bis 7.36 weisen die Ergebnisse der multiplen Korrelationsanalysen in den Klassen 1 bis 36 aus. Die Reihenfolge der fünf Urteilsdimensionen und deren Beiträge zur Aufklärung der Notenvarianzen fallen für die einzelnen Lehrer unterschiedlich aus und variieren bei einzelnen Lehrern, je nachdem, ob Jungen oder Mädchen beurteilt werden.

Die Ergebnisse der multiplen Korrelationsanalysen von fünf Lehrern wurden herausgegriffen und interpretativ beschrieben.

6. Diskussion der Ergebnisse

Aus der Vielfalt der Einstellungen und Erwartungen bei Lehrern und ihren Auswirkungen auf die Beurteilung von Schülern haben wir uns, in Anlehnung an die Befunde HOFERs (1969), der Frage zugewandt, wie Lehrer ihre implizite Persönlichkeitstheorie zur aktuellen Schülerbeurteilung im einzelnen in ihren Klassen realisieren. Damit treten in unserer Arbeit individuelle Unterschiede des Lehrers bei der Schülerbeurteilung in den Blickpunkt des Interesses. Zugleich liegt der Schwerpunkt dieser Arbeit auf der Entwicklung und Erprobung eines Instrumentariums, das dem einzelnen Lehrer eine deskriptive Analyse seines Urteilsverhaltens Schülern gegenüber ermöglichen soll.

Die deskriptive Analyse der Beurteilungen gibt keine Auskunft über die "Richtigkeit" von Lehrerurteilen über Schüler, denn Beurteilungen stellen einen Teilaspekt, die kognitive und konative Komponente sozialen Verhaltens, dar. Beurteilungen lassen sich daher nicht loslösen vom

übergeordneten (komplexeren) Forschungsgegenstand der sozialen Interaktion, sondern wären für sich allein gesehen nur ein Teil, der auf einem niedrigeren Komplexitätsniveau analysiert wird, wie es jahrzehntelang der Zugang zum angesprochenen Problem war (vergl. MERZ, 1962). Doch dieser Zugang ist unzweckmäßig, da "das Bild vom Anderen" (KAMINSKI, 1959) selbst zum sozialen Verhalten gehört. Dabei wird die Beurteilung betrachtet als abhängig von:

(1) der Eigenart des Beobachteten;
(2) der Eigenart des Beobachters;
(3) der Eigenart der Information, die der Beobachter über den Beobachteten bekommen hat;
(4) der Eigenart des situativen Kontextes, in dem sich Beobachter und Beobachteter befinden.

Da Relationen zwischen diesen Punkten betrachtet werden, wird die Frage nach der "Richtigkeit" der Urteile abgelöst von der Frage: Wie spielt sich der Prozeß der Urteilsbildung ab?

Für unsere Gesamtstichprobe von N=810 beurteilten Schülern konnten die von HOFER (1969) gefundenen Dimensionen zur aktuellen Schülerbeurteilung repliziert werden. Darüberhinaus ließen sich für alle an der Untersuchung beteiligten Klassen bzw. Lehrern in der Regel drei unterscheidbare, clusteranalytisch ermittelte Schülergruppen nachweisen, die aufgrund ihrer Ähnlichkeit ihrer vom Lehrer erhaltenen Beurteilungsscores auf den fünf HOFERschen Urteilsdimensionen zusammengefaßt werden konnten. Allerdings sagt diese konsistente Dreiergruppierung je Klasse bzw. Lehrer allein noch nichts darüber aus, wie Lehrer im einzelnen das implizite Persönlichkeitsmodell zur aktuellen Schülereinschätzung realisieren, d.h. durch welche Eigenschaften

unterschiedlich gruppierte Schüler einer Klasse nach Ansicht ihres Lehrers charakterisiert sind. Um die beim einzelnen Lehrer ermittelten und als Beurteilertendenzen zu interpretierenden Schülergruppen besser beschreiben zu können, stellen wir die Schülergruppierungen pro Klasse als Gruppenprofile dar, die dem einzelnen Lehrer zugleich eine anschauliche Rückmeldung bezüglich seiner Urteilstendenzen ermöglichen sollen. Hierbei stand nicht die Frage nach der "Richtigkeit" (Validitätsaspekt) beurteilter Schülergruppen im Vordergrund, sondern vielmehr die Frage, wie sich unter Zugrundelegung des impliziten Persönlichkeitsmodells zur aktuellen Schülerbeurteilung der Prozeß der Urteilsbildung beim einzelnen Lehrer abspielt. Als Kriterien zur Interpretation der Gruppenprofile bei einzelnen Lehrern wurden verwendet:

(1) Die SCHEFFÉ-Methode zur Auswahl signifikant trennender Urteilsdimensionen;

(2) Die signifikante Abweichung eines Gruppenmittelwertes vom Klassenmittelwert, die durch das Vertrauensintervall (CL) über den Standardmeßfehler definiert ist.

(3) Die Dimensionen Disziplintiertheit vs. Undisziplintiertheit und geistige Regsamkeit als Bewertungskriterien im Sinne von positiv - negativ.

Zur Beurteilung der Schüler ihrer Klassen haben wir den Lehrern die von HOFER (1969) festgelegte Liste von 25 polaren Eigenschaftspaaren vorgelegt. Wenn wir einräumen, daß diese Liste eine repräsentative Auswahl von Eigenschaften darstellt, die Lehrer bei der Beurteilung von Schülern benutzen, so läßt sich nicht von vornherein der Einwand entkräften, daß die von HOFER selegierten Eigenschaftsbegriffe die Lehrer zu schematisierten Beurteilungen herausfordern und eine andere Eigenschaftsauswahl möglicherweise zu anderen Beurteilungsdimensionen und anderen Beur-

teilertendenzen bei einzelnen Lehrern hätte führen können. Dieser Einwand verliert aber an Argumentationswert, wenn wir die hohen Korrelationen zwischen den HOFER-Faktoren und den Zeugnisnoten in verschiedenen Fächern bei den einzelnen Lehrern in Rechnung stellen. Wenngleich die HOFER-Faktoren als Prädiktoren in Reihenfolge und Beitrag zur Varianzaufklärung der Noten von Lehrer zu Lehrer beträchtlich variieren, so bleiben doch die hohen Korrelationen zwischen den Prädiktoren und den Kriterien insgesamt erhalten. Zudem zeichnet sich ein Trend ab, daß bei Jungen insgesamt mehr Notenvarianz durch die HOFER-Faktoren aufgeklärt wird als bei Mädchen. Hier wird deutlich, daß sich das Konstrukt "implizite Persönlichkeitstheorie" in bezug auf die Zeugnisnoten als valide im Sinne der concurrent-validity erweist.

An dieser Stelle sei noch einmal auf die Höhe der Validitätskoeffizienten in der Größenordnung zwischen 0.7 und 0.8 hingewiesen. Obgleich derartige Koeffizienten nach LIENERT (1967[2], S. 310) als hoch bis sehr hoch zu bezeichnen sind, werden durchschnittlich nur 50-65% der Notenvarianz durch die HOFER-Faktoren aufgeklärt. Betrachten wir jedoch die HOFER-Faktoren unter dem Gesichtspunkt der Zulänglichkeit (vergl. LIENERT 1967[2], S. 17), so erhalten wir Aufschluß darüber, inwieweit die HOFER-Faktoren für die Zeugnisnoten tatsächlich repräsentativ sind. Das Konzept der Zulänglichkeit schließt den Aspekt der Kriteriumsreliabilität, in unserem Falle die Reliabilität der Noten, die wir in unserer Untersuchung nicht überprüft haben, mit ein. Die Validität des prognostischen Instrumentariums "HOFER-Faktoren" ist dann sowohl auf seine eigene Reliabilität als auch auf die der Noten zu beziehen.

Nach TENT (1969, S. 134) ergibt sich die Zulänglichkeit aus dem Verhältnis der aufgeklärten zu der überhaupt

aufklärbaren Varianz: $r^2_{tc}/(r^2_{tt'} r^2_{cc})$. TENT (1969, S. 134) führt hierzu aus: "Auf diese Weise ist es möglich, empirisch gefundene Validitäten zu beurteilen und miteinander zu vergleichen. Setzt man für die maximal als gemeinsam aufklärbare Varianz das Mittel der beiden Determinationskoeffizienten, erhält man für die Zulänglichkeit

$$r^2_{t'c'} = \frac{r^2_{tc}}{\sqrt{r^2_{tt'} \; r^2_{cc}}} \quad ; \quad \text{bzw. } r_{t'c'} = \frac{r_{tc}}{\sqrt{r_{tt'} \; r_{cc}}} \; ,$$

also den doppelt minderungskorrigierten Validitätskoeffizienten S P E A R M A N S (LIENERT: crit. et test corr. r_{tc}, 1967, S. 298). Dieser Ausdruck ist, so verstanden, natürlich kein empirisch interpretierbarer eignungsdiagnostischer Validitätskoeffizient. Er ist vielmehr ein Index für die relative Güte der Validität eines prognostischen Instruments in einer Prädiktor-Kriteriums-Konstellation".

Aus dem oben erläuterten doppelt minderungskorrigierten Validitätskoeffizienten läßt sich errechnen, inwieweit die fünf Beurteilungsdimensionen HOFERs für die Zeugnisnoten zulänglich sind. Dabei legen wir als Maß für die Enge des Zusammenhanges zwischen den Zeugnisnoten und den HOFER-Faktoren ein durchschnittliches R von 0,70 (vergl. Tab. 3 und für die fünf HOFER-Faktoren ein mittleres r_{stab} von 0,80 zugrunde (vergl. Tab. 2). Der schulischen Leistungsbeurteilung (den Noten) kommt nach TENT (1969, S. 134) ein mittleres r_{cc} von 0,60 zu. Hieraus errechnet sich eine Zulänglichkeit von

$$r_{t'c'} = \frac{0,70}{\sqrt{0,80 \cdot 0,60}}$$

Diesen Überlegungen zufolge sind die fünf Beurteilungsdimensionen HOFERs repräsentativ für die Zeugnisnoten der an dieser Untersuchung beteiligten Lehrer.

L i t e r a t u r v e r z e i c h n i s

ADAMS, H. F.: The good judge of personality. J. abnorm. soc. Psychol. 1927, 22, 172 - 181.

ADINOLFI, A. A.: Relevance of person perception research to clinical psychology. J. consult. clin. Psychol. 1971, 37, 167 - 176.

ALLPORT, G. W.: Personality. New York 1937.

ARLITT, H.: Variability among a group of judges in rating character traits in children. Psychol. Bull. 1926, 23, 617 - 619.

ASCH, S. E.: Forming impressions of personality. J. abnorm. soc. Psychol. 1946, 41, 258 - 290.

ASH, P.: The reliability of psychiatric diagnosis. J. abnorm. soc. Psychol. 1949, 44, 272 - 276.

BARTRAM, M., ROLLETT, B. & S. SUHRE: Kriterien der Schülerbeurteilung von Lehrern. Unveröffentlichte Untersuchung an der PH Niedersachsen, Abt. Osnabrück 1971.

BECKER, W. C.: The relationship of factors in parental ratings of self and each other to the behavior of Kindergarten children as rated by mothers, fathers and teachers. J. consult. Psychol. 1960, 24, 507 - 527.

BENDER, J. E. & A. H. HASTORF: The perception of persons: Forecasting another person's responses on three personality scales. J. abnorm. soc. Psychol. 1950, 45, 556 - 561.

BIERI, J.: Cognitive complexity-simplicity and predictive behavior. J. abnorm. soc. Psychol. 1955, 51, 263 - 268.

BIERI, J., ATKINS, A. L., BRIAR, A. S., LEAMAN, R. L., MILLER, H. & T. TRIPOLDI: Clinical and social judgement: The discrimination of behavioral information. New York 1966.

BRUNER, J. S., SHAPIRO, D. & R. TAGUIRI: The meaning of traits in isolation and in combination. In R. TAGUIRI & L. PETRULLO (Eds.), Person perception and interpersonal behavior. Stanford 1958.

BRUNSWIK, E.: Social perception of traits from photographs. Psychol. Bull. 1945, 42, 535 - 536.

BUSS, A. H., WIENER, M., DURKEE, A. & B. BAER: The measurement of anxiety in clinical situations. J. consult. Psychol. 1955, 19, 125 - 129.

CAREY, J. F., BERG, J. A. & A. C. VAN DUSEN: Reliability of ratings of employee satisfaction based on written interview records. J. appl. Psychol. 35, 252 - 255.

CENTERS, R.: An effective classroom demonstration of stereotypes. J. soc. Psychol. 1951, 34, 41 - 46.

CHANCE, J. E. & W. MEADERS: Needs and interpersonal perception. J. Person. 1960, 28, 200 - 209.

CHAREN, S.: The awareness of hostile feeling in patients by their nurses. J. consult. Psychol.1955, 19, 290.

CROCKETT, W. H. & T. MEIDINGER: Authoritarianism and interpersonal perception. J. abnorm. soc. Psychol. 1956, 53, 378 - 380.

CRONBACH, L. J.: Proposals leading to analytical treatment of social perception scores. In: TAGUIRI, R. & PETRULLO, L. (Hrsg.): Person perception and interpersonal behavior. Stanford 1958, 353 - 379.

DUBIN, S. S., BURKE, L. K., KATZ, A. & D. J. CHESLER: Characteristics of rates whose ratings agree with consensus ratings. USA Personn. Res. Br. Note, 1954, No 35, 8 p.

DYMOND, R. F.: A preliminary investigation of the relation of insight and empathy. J. consult. Psychol. 1948, 12, 228 - 233

DYMOND, R. F.: Can clinicans predict individual behavior? J. Pers. 1953/54, 22, 151 - 161.

EARLE, M. G.: The relation between personality and character traits and intelligence. J. appl. Psychol. 1926, 10, 453 - 461.

ERTEL, S.: Kategorien der Personwahrnehmung und ihre kulturelle Bedingtheit. Psychol. Forsch. 1964, 27, 475 - 540.

FAY, P. J. & W. C. MIDDLETON: Judgement of occupation from voice transmitted over a public address system and over radio. J. appl. Psychol. 1929, 23, 586 - 601.

FLANAGAN, J. C.: The critical incidence technique. Psychol. Bull. 1954, 51, 327 - 358.

GAGE, N. L.: Handbuch der Unterrichtsforschung Teil I. Deutsche Bearbeitung des "Handbook of Research on Teaching". Hrsg. Ingenkamp K.: Weinheim, Berlin, Basel 1970.

GEBAUER, T.: Vergleichende Untersuchung über den Voraussagewert von Aufnahmeprüfung und Testprüfung für den Erfolg auf der höheren Schule. Nat. Diss. Köln 1964

GORDON, J. E.: Interpersonal predictions of repressors and sensitisers J. Pers. 1956/57, 25, 686 - 698.

GUILFORD, J. P.: Psychometric methods. New York 1954. 278, 281.

HALPERN, H. M.: Empathy, similarity and self-satisfaction. J. consult. Psychol. 1955, 19, 449 - 452.

HARVEY, O., HUNT, D. E. & H. M. SCHRODER: Conceptional systems and personality organisation. New York, Wiley 1961

HELPER, M. M.: Parental evaluation of children and children's self-evaluation. J. abnorm. soc. Psychol. 1958, 56, 190 - 194.

HEYMANS, G. & E. WIERSMA: Beiträge zur speziellen Psychologie aufgrund einer Massenuntersuchung. Z. Psychol. 1906, 42, 81 - 127.

HITPASS, J.: Bericht über eine sechsjährige Bewährungs-Kontrolle von Aufnahmeprüfung und Testprüfung. Schule & Psychol. 10, 211 - 218, 1963.

HOFER, M.: Die Schülerpersönlichkeit im Urteil des Lehrers. Weinheim 1969.

HOFSTÄTTER, P. R.: Psychologie. Frankfurt 1957.

HOLLANDER, E. P.: Studies of leadership among Naval Aviation Cadets. J. aviat. Med., 1954, 25, 164 - 170, 200.

INGENKAMP, K.: Untersuchungen zum Übergang auf weiterführende Schulen nach der 6. Klasse. In: INGENKAMP, K.: Pädagogisch-psycholog. Untersuchungen zum Übergang auf weiterführende Schulen. Weinheim Beltz 1963.

JURGENSON, C. E.: Overall Job Success as a Basis for Employee Rating. J. Appl. Psychol. 1950, 34, 333 - 337.

KAMINSKI, G.: Das Bild vom Anderen. Berlin 1959.

KELLY, G. A.: The Psychology of Personal Constructs. New York 1955.

KESSEN, W.: The Role of Experience in Judging Children's Photographs. J. abnorm. Psychol. 1957, 54, 375 - 379.

KNIGHT, F. B.: The Effect of the "Acquaintance Factor" upon Personal Judgments. J. educ. Psychol. 1923, 14, 129 - 142.

KROHNE, H. W.: Psychologischer Stress, Angstkontrolle und Differenziertheit der Personwahrnehmung. Z. f. Sozialpsychol. 1973, Bd. 4, Heft 2, S. 88.

LÄMMERMANN, H.: Das Mannheimer kombinierte Verfahren der Begabtenauslese. Leipzig 1927.

LAING, G.: Group interaction and perceived similarity of members. J. abnorm. soc. Psychol. 1960, 61, 335 - 340.

LEVENTHAL, H.: Cognitive Processes and Interpersonal Prediction. J. abnorm. soc. Psychol. 1957, 55, 176 - 180.

LEVIN, H.: The Influence of Fullness of Interview on the Reliability, Discriminability and Validity of Interview Judgements. J. consult. Psychol. 1954, 18, 303 - 306.

LIENERT, G. A.: Testaufbau und Testanalyse. Weinheim 1967^2.

LINDZEY, G.: Handbook of Social Psychology. Cambridge, Mass. 1954.

LITTERER, O. F.: Stereotypes. J. soc. Psychol. 1933, 4, 59 - 69.

LUFT, J.: Implicit Hypotheses and Clinical Predictions. J. abnorm. soc. Psychol. 1950, 45, 756 - 759.

LUNDY, R. M.: Assimilative Projection and Accuracy Prediction in Interpersonal Perceptions. J. abnorm. soc. Psychol. 1956, 52, 33 -

MASON, D. J.: Judgements of Leadership Based upon Physionomic Cues. J. abnorm soc. Psychol. 1957, 54, 273 - 274.

Mc NEMAR, Q.: Psychological Statistics. New York, London, Wiley, 196

MERZ, F.: Über die Beurteilung der persönlichen Eigenart unserer Mitmenschen. Arch. ges. Psychol. 1962, 114, 187 - 221.

MITTENECKER, E.: Die Variation von Lern- und Verstärkungsbedingungen bei der eindrucksmäßigen Beurteilung von Persönlichkeitsmerkmalen. Ber. 22. Kongr. DGfB 1959, Göttingen 1960, 62 - 64.

MÜLLER, R. G. E.: Verhaltensstörungen bei Schulkindern, München 1970.

NORMAN, R. D. & P. AINSWORTH: The Relationship among Projection, Empathy and Adjustment, Operationaly Defined. J. consult. Psychol. 1954, 18, 53 - 58.

NOSSBERGER-EIDLER, F.: Intelligenzschätzung nach dem Gesichtsausdruck. Psychol. u. Praxis. 3, 281 - 289. 1959.

OLSON, W. C. & N. M. WILKINSON: Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior control. Educ. Admin. Superv. 1938, 24, 81 - 93.

PRECKER, J. A.: The automorphic process in the attribution of values. J. Pers. 1952, 21, 356 - 363.

RANK, Therese: Schulleistungen und Persönlichkeit. München 1962.

RASCH, W. D.: Hat sich die Graphologie bewährt? Bern 1957.

REMMERS, H. H.: Reliability and halo-effect of high school and college students' judgements of their teachers. J. appl. Psychol. 1934, 18, 619 - 630.

RICE, S. A.: Stereotypes: A source of error in judging human character. J. Pers. Research 1927, 5, 267 - 276.

SACHS, L.: Statistische Auswertungsmethoden. Berlin, Heidelberg, New York 1969.

ORLIK, P.: Kritische Untersuchungen zur Begabtenförderung. Meisenheim, 1967.

SARBIN, T. R. TAFT, R. & D. E. BALLEY: Clinical inference and cognitive theory. New York: Holt, Rinehart & Wiston, 1960.

SCHRODER, H. M., DRIVER, M. J. & S. Streufert: Human information processing: Individuals and groups functioning in complex social situations. New York 1967.

SCHULTZE, W.: Über den Voraussagewert der Auslese - Kriterien für den Schulerfolg am Gymnasium, Max-Traeger-Stiftung; Forschungsberichte 1. Frankfurt, 1964

SCODEL, A. & M. L. FREEDMAN: Additional observations on the social perceptions of authoritarians and nonauthoritarians. J. abnorm. soc. Psychol. 1956, 52, 92 - 95.

SEARS, R. R.: Experimental studies of projection. J. soc. Psychol. 1936, 7, 151 - 163.

SIEGEL, A. J.: An experimental evaluation of the sensitivity of the empathy test. J. appl. Psychol. 1954, 38, 222 - 223.

SMITH, A. J.: The attribution of similarity: The influence of success and failure. J. abnorm. soc. Psychol. 1960, 61, 419 - 423.

SNOW, M. J.: An experiment in the validity of judging human ability. J. appl. Psychol. 1924, 8, 339 - 346.

SPANNER, M.: Attribution of traits and emotional health as factors associated with the prediction of personality characteristics of others. J. consult. Psychol. 1961, 25, 210 - 215.

STEINKAMP, G.: Die Rolle des Volksschullehrers im schulischen Selektionsprozess, Hamburger Jahrbuch f. Wirtsch.- u. Gesellsch. Politik, 12, 1967.

STOCKFORD, L. & H. W. BISSELL: Factors involved in establishing a ment-rating scale. Personnel, 1949, 26, 94 - 116.

STONE, P. & J. KAMIYA: Judgments of consensus during group discussion. J. abnorm. soc. Psychol. 1957, 55, 171 - 175.

STROBL, S.: Eine Untersuchung über die Brille als Ausdruckfaktor. Z. Psychol. 1956, 159, 101 - 129.

TAFT, R.: The ability to judge people. Psychol. Bull. 1955, 52, 1 - 23.

TAFT, R.: Some characteristics of good judges of others. Brit. J. Psychol. 1956, 47, 19 - 29.

TENT, L.: Die Auslese von Schülern für weiterführende Schulen. Göttingen 1969[2].

TENT, L. & H. REMMERS: Schätzverfahren in der Unterrichtsforschung. In: K. INGENKAMP (Hrsg.): Handbuch der Unterrichtsforschung I, Weinheim, Berlin, Basel 1970.

THORNDIKE, E. L.: A constant error in psychological ratings. J. appl. Psychol. 1920, 4, 25 - 90.

THORNTON, G. R.: The effect upon judgments of personality traits of varying a single factor in photograph. soc. Psychol. 1943, 18, 127 - 148.

TORGERSON, W. S.: Theory and Methods of Scaling. New York, London, Sidney 1958[7].

ÜBERLA , K.: Faktorenanalyse, Berlin, Heidelberg, New York; Springer 1968.

VELDMAN, D. J.: Fortran Programming for the Behavioral Sciences. Holt, Rinehart & Winston, New York 1968, 308 - 317.

VERNON, P. E.: Some characteristics of the good judge of personality. J. soc. Psychol. 1933, 4, 42 - 58.

WARD, J. H.: Hierarchical grouping analysis to optimize an objective function. J. Am. Statist. Assoc. 58, 236 - 244.

WEINGARDT, E.: Korrelation und Voraussagewert von Zeugnisnoten bei Gymnasiasten. Erziehung und Psychologie Nr. 3, München, Basel

WEINGARTEN, E. M.: A study of selective perception in clinical judgment. J. Pers. 1948/49, 17, 369 - 406.

A n h a n g

Instruktion zum Ausfüllen des Polaritätenprofils

Im folgenden werden Ihnen 25 Eigenschaftswörter und deren Gegensätze (z. B. freundlich - unfreundlich) vorgelegt, mit deren Hilfe Sie bitte jeden Schüler bzw. jede Schülerin in Ihrer Klasse charakterisieren wollen.

Zwischen den Eigenschaften befindet sich eine 7-Punkteskala. Mit Hilfe dieser Skala können Sie die Ausprägung einer der beiden Eigenschaften bei dem zu beurteilenden Schüler angeben.

Bei der Beurteilung rufen Sie sich bitte erst den betreffenden Schüler (z. B. Sepp Maier) ins Gedächnis.
Sodann überlegen Sie bitte, welche der beiden Eigenschaften (z. B. aufmerksam oder abgelenkt) den Schüler besser beschreibt. Dann bestimmen Sie bitte noch die Ausprägung dieser Eigenschaften. Dabei bedeutet:
Der Punktwert 3 eine starke, der Punktwert 2 eine mittlere und der Punktwert 1 eine schwache Ausprägung. Der Punktwert 0 bedeutet, daß keine der beiden Eigenschaften zutrifft.

Halten Sie beispielsweise der Schüler Sepp Maier eher für aufmerksam als für abgelenkt, dann durchkreuzen Sie bitte eine der drei Klammern auf der linken Seite der Skala und zwar:
Ist der Schüler sehr aufmerksam, dann durchkreuzen Sie die Klammer mit der 3, bei mittlerer Aufmerksamkeit die Klammer mit der 2, bei schwacher Aufmerksamkeit die Klammer mit der 1 (jeweils auf der linken Seite der Skala). Halten Sie den Schüler eher für abgelenkt, dann durchkreuzen Sie bitte in analoger Weise eine der drei Klammern auf der rechten Seite, halten Sie den Schüler weder für aufmerksam noch für abgelenkt, dann durchkreuzen Sie bitte die Klammer mit der 0.

Nun noch eines. Bei der Beurteilung der Schüler durch solche Polaritätenprofile kann man die Einschätzung aufgrund sehr unterschiedlich langer und sorgfältiger Überlegungen abgeben.

Außerdem wechselt das Verhalten der Schüler in unterschiedlichen Situationen und damit auch die Beurteilung. Stellen Sie sich bitte das typische (häufigste) Verhalten des Schülers im Klassenverband vor und geben Sie dann ruhig ein intuitives Urteil ab. Das heißt, überlegen Sie bei der Beurteilung nicht zu lange und geben Sie dem ersten Eindruck nach.

Name und Vorname des Schülers: SEPP MAIER

1.	uninteressiert	(3)	(2)	(1)	(0)	(1)	(2)	(3)	interessiert
2.	verschlossen	(3)	(2)	(1)	(0)	(1)	(2)	(3)	offen
3.	sensibel	(3)	(2)	(1)	(0)	(1)	(2)	(3)	robust
4.	selbstsicher	(3)	(2)	(1)	(0)	(1)	(2)	(3)	selbstunsich
5.	unbegabt	(3)	(2)	(1)	(0)	(1)	(2)	(1)	begabt
6.	einfallsreich	(3)	(2)	(1)	(0)	(1)	(2)	(3)	einfallslos
7.	gesellig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	ungesellig
8.	geltungsbedürftig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	bescheiden
9.	fleißig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	faul
10.	kompliziert	(3)	(2)	(1)	(0)	(1)	(2)	(3)	unkomplizier
11.	intelligent	(3)	(2)	(1)	(0)	(1)	(2)	(3)	dumm
12.	ehrgeizig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	gleichgültig
13.	unkonzentriert	(3)	(2)	(1)	(0)	(1)	(2)	(3)	konzentriert
14.	höflich	(3)	(2)	(1)	(0)	(1)	(2)	(3)	frech
15.	unordentlich	(3)	(2)	(1)	(0)	(1)	(2)	(3)	ordentlich
16.	schüchtern	(3)	(2)	(1)	(0)	(1)	(2)	(3)	dreist
17.	folgsam	(3)	(2)	(1)	(0)	(1)	(2)	(3)	unfolgsam
18.	Unruhig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	ruhig
19.	pflichtbewußt	(3)	(2)	(1)	(0)	(1)	(2)	(3)	pflichtverg
20.	aufmerksam	(3)	(2)	(1)	(0)	(1)	(2)	(3)	abgelenkt
21.	aufgeweckt	(3)	(2)	(1)	(0)	(1)	(2)	(3)	träge
22.	ausgelichen	(3)	(2)	(1)	(0)	(1)	(2)	(3)	launisch
23.	zuverlässig	(3)	(2)	(1)	(0)	(1)	(2)	(3)	unzuverläss
24.	sympatisch	(3)	(2)	(1)	(0)	(1)	(2)	(3)	unsympatisc
25.	führend	(3)	(2)	(1)	(0)	(1)	(2)	(3)	zurückhalte

Tabellen 1 bis 36

Clusteranalytisch ermittelte Schülergruppierungen pro Klasse

Legende

$\bar{X}_{ij}$ = Gruppenmittelwerte auf den fünf HOFERschen Beurteilungsdimensionen, wobei der Index i die Gruppenzugehörigkeit angibt; j bezeichnet die Dimensionszugehörigkeit.

T_{ij} = T-transformierter $\bar{X}_{ji}$- Wert

$\bar{\bar{X}}_{j}$ = Klassenmittelwert bezüglich der fünf Beurteilungsdimensionen

s_{j} = Standardabweichung in einer Klasse bezüglich der fünf Beurteilungsdimensionen

r_{stab} = Stabilitätskoeffizient

Gruppe
1-2
1-3
2-3
} Multiple Mittelwertsvergleiche zwischen den jeweils drei Gruppen pro Klasse. Signifikante Mittelwertsdifferenzen auf den Dimensionen sind durch Sterne gekennzeichnet.

Kl. 1

Schüler Nr.	I	II	III	IV	V	
012	26	25	20	10	3	
011	27	34	15	8	4	
009	26	26	16	7	7	Gruppe 1
004	16	28	11	5	2	
003	27	34	18	9	3	
$\bar{X}_1$	24.4	29.4	16.0	7.8	3.8	
T_1	47	61	60	42	42	
010	23	19	10	13	10	
002	31	18	13	14	10	Gruppe 2
$\bar{X}_2$	27.0	18.5	11.5	13.5	10.0	
T_2	54	47	48	62	61	
008	24	13	9	8	6	
007	27	18	10	10	5	
006	23	10	8	12	12	Gruppe 3
005	24	11	9	10	10	
001	32	16	8	15	6	
$\bar{X}_3$	26.0	13.6	8.8	11.0	7.8	
T_3	51	41	41	53	54	
$\bar{\bar{X}}$	25.5	21.0	12.3	10.1	6.5	
s	3.9	7.9	3.8	2.8	3.1	
r_{stab}	.66	.92	.88	.55	.87	
Gruppe						
1 - 2		*		*	*	
1 - 3		*	*			
2 - 3						

Kl. 2

hüler	I	II	III	IV	V	
6	13	25	7	5	8	
4	23	25	12	7	9	
0	15	23	9	5	9	
9	13	28	12	5	8	Gruppe 1
8	19	32	14	4	7	
7	17	23	12	8	9	
4	17	35	7	5	8	
$\bar{X}_1$	16.7	27.3	10.4	5.6	8.3	
T_1	42	58	51	45	47	
5	24	15	11	8	11	
3	22	23	13	11	9	Gruppe 2
2	21	9	12	5	11	
5	25	14	6	4	10	
$\bar{X}_2$	23.6	15.3	10.5	7.0	10.3	
T_2	53	42	51	49	62	
1	28	16	9	8	9	
6	35	13	12	17	8	Gruppe 3
3	25	17	7	9	6	
$\bar{X}_3$	29.3	15.3	9.3	11.3	7.7	
T_3	64	42	46	62	42	
$\bar{X}$	21.2	21.3	10.2	7.2	8.7	
s	5.9	7.3	2.5	3.4	1.3	
r_{stab}	.78	.81	.74	.74	.29	
uppe						
- 2	*	*			*	
- 3	*	*		*		
- 3					*	

Kl. 3

Schüler Nr.	I	II	III	IV	V	
044	8	20	9	13	8	
041	10	28	10	8	10	
035	13	19	7	12	10	Gruppe 1
033	12	26	10	9	10	
030	16	30	9	8	7	
$\bar{X}_1$	11.8	24.6	9.0	10.0	9.0	
T_1	39	44	44	42	57	
043	33	21	21	20	12	
038	50	21	17	28	4	
037	41	25	15	26	4	Gruppe 2
032	44	40	14	20	5	
$\bar{X}_2$	42	26.8	16.8	23.5	6.3	
T_2	65	47	65	67	46	
042	33	37	12	13	8	
040	23	38	16	11	4	
039	26	23	9	15	8	
036	24	31	12	12	7	
034	27	32	15	12	9	Gruppe 3
031	14	31	8	11	6	
029	19	25	9	12	6	
028	25	35	11	15	4	
027	24	30	9	12	9	
$\bar{X}_3$	23.9	31.3	11.2	12.6	6.8	
T_3	49	55	50	47	48	
$\bar{\bar{X}}$	24.6	28.4	11.3	14.3	7.3	
s	11.6	6.2	3.6	5.5	2.4	
r_{stab}	.94	.62	.78	.92	.57	
Gruppe						
1 - 2	*		*	*		
1 - 3	*					
2 - 3	*		*	*		

Kl. 4

Schüler Nr.	I	II	III	IV	V	
158	40	37	24	14	3	
157	38	34	15	17	2	
154	28	29	16	14	7	
153	36	34	19	13	5	
152	35	26	19	15	5	Gruppe 1
150	28	23	18	14	3	
148	34	31	20	14	7	
146	37	41	24	11	2	
145	30	22	17	17	8	
$\bar{X}_1$	34.0	30.8	19.1	14.3	4.7	
T_1	50	53	51	48	50	
156	49	25	23	21	4	
155	45	17	25	25	2	Gruppe 2
147	46	23	23	21	6	
$\bar{X}_2$	46.7	21.7	23.7	22.3	4.0	
T_2	63	42	60	65	47	
151	23	15	8	11	8	
149	12	38	9	6	2	Gruppe 3
$\bar{X}_3$	17.5	26.5	8.5	8.5	5.0	
T_3	32	48	30	35	62	
$\bar{X}$	34.4	28.2	18.6	15.2	4.6	
s	9.4	7.7	5.1	4.6	2.2	
r_{stab}	.92	.82	.89	.91	.76	

Gruppe					
1 - 2	*			*	
1 - 3	*		*	*	
2 - 3	*		*	*	

Kl. 5

Schüler Nr.	I	II	III	IV	V	
075	21	19	15	7	11	
070	27	13	12	19	9	
068	33	15	23	16	10	
067	27	17	14	15	8	
066	26	24	15	10	4	Gruppe 1
064	32	13	14	19	7	
060	35	19	14	10	11	
059	32	10	19	16	10	
$\bar{X}_1$	29.1	16.3	15.8	12.8	8.8	
T_1	53	42	45	54	53	
074	41	29	23	8	5	
073	23	41	20	9	7	
071	32	26	24	15	6	Gruppe 2
069	22	39	24	6	5	
063	34	37	21	9	4	
$\bar{X}_2$	30.4	34.4	22.4	9.4	5.4	
T_2	55	62	62	47	39	
072	19	19	15	7	9	
065	19	25	18	7	12	
062	16	21	15	6	9	Gruppe 3
061	14	33	19	4	11	
$\bar{X}_3$	17.0	24.5	16.8	6.0	10.3	
T_3	37	51	47	40	59	
$\bar{X}$	26.6	23.5	17.9	10.8	8.1	
s	7.3	9.2	3.9	4.7	2.5	
r_{stab}	.78	.88	.83	.84	.31	

Gruppe					
1 - 2		*	*		*
1 - 3	*			*	
2 - 3	*		*		*

Kl. 6

chüler r.	I	II	III	IV	V	
185	13	44	21	4	6	
184	17	34	14	12	3	
183	10	37	11	5	6	Gruppe 1
182	26	37	15	7	7	
$\bar{X}_1$	16.5	38.0	15.3	7.0	5.5	
T_1	38	60	47	41	46	
181	43	22	22	15	6	
179	49	30	25	12	5	
178	38	18	17	14	5	Gruppe 2
177	47	27	19	11	9	
$\bar{X}_2$	44.3	24.3	20.8	13.0	6.3	
T_2	51	46	58	55	51	
180	46	15	9	10	8	
176	42	19	16	19	7	Gruppe 3
$\bar{X}_3$	44.0	17.0	12.5	14.5	7.5	
T_3	58	38	41	58	58	
$\bar{X}$	33.1	28.3	16.9	10.9	6.2	
s	14.4	9.2	4.7	4.4	1.6	
r_{stab}	.95	.86	.87	.82	.64	

ruppe		
1 - 2	*	*
1 - 3	*	*
2 - 3		

Kl. 7

Schüler Nr.	I	II	III	IV	V	
100	42	23	16	14	6	
097	38	26	14	26	5	Gruppe 1
096	39	18	16	19	8	
$\bar{X}_1$	39.7	19.0	15.3	16.3	6.3	
T_1	63	40	48	59	46	
099	20	19	9	12	10	
098	20	27	16	13	7	
092	14	21	11	9	5	
091	24	19	15	9	12	Gruppe 2
090	19	34	14	6	8	
088	18	26	13	7	9	
087	24	18	14	9	11	
$\bar{X}_2$	19.9	23.4	13.1	9.3	8.9	
T_2	43	45	43	45	57	
095	46	34	22	15	5	
094	28	28	25	9	7	
093	23	46	23	6	6	Gruppe 3
089	28	23	21	14	7	
086	17	45	15	8	4	
$\bar{X}_3$	28.4	35.2	21.2	10.4	5.8	
T_3	52	58	61	47	43	
$\bar{X}$	26.7	27.8	16.3	11.7	7.3	
s	9.7	8.8	4.4	5.2	2.3	
r_{stab}	.94	.87	.89	.88	.50	

Gruppe					
1 – 2	*			*	
1 – 3			*	*	
2 – 3			*		

K. 8

Schüler Nr.	I	II	III	IV	V	
115	30	23	17	7	10	
113	22	37	14	10	6	
112	37	35	19	9	5	
111	26	36	13	10	8	
109	36	33	15	9	4	Gruppe 1
104	36	39	18	8	4	
103	24	48	19	8	3	
102	38	45	21	11	6	
101	28	42	18	8	8	
$\bar{X}_1$	30.8	37.6	17.1	8.9	6.0	
T_1	48	56	56	45	49	
114	41	15	15	19	13	
110	44	14	9	22	7	
108	42	18	10	14	2	Gruppe 2
105	48	22	16	14	7	
$\bar{X}_2$	43.8	17.3	12.5	17.3	7.3	
T_2	62	37	43	60	53	
107	17	23	10	6	6	
106	20	29	13	8	7	Gruppe 3
$\bar{X}_3$	18.5	26.0	11.5	7.0	6.5	
T_3	35	45	40	42	50	
$\bar{X}$	32.6	31.2	15.1	11.5	6.4	
s	9.2	10.4	3.5	5.5	2.7	
r_{stab}	.83	.94	.60	.87	.54	

Gruppe				
1 - 2	*	*	*	
1 - 3	*			
2 - 3	*		*	

KL. 9

Lauser Nr.	I	II	III	IV	V	
129	26	39	19	9	13	
128	14	45	12	4	7	
126	32	36	19	12	7	
124	18	32	17	8	12	Gruppe 1
121	28	42	11	11	7	
120	31	41	23	8	8	
119	23	33	9	15	5	
$\bar{X}_1$	24.6	38.3	15.7	9.6	8.4	
T_1	47	58	52	47	49	
127	41	18	18	20	12	
125	34	28	16	16	9	
117	38	29	19	13	5	Gruppe 2
116	43	15	13	17	9	
$\bar{X}_2$	39.0	22.5	17.0	16.5	8.8	
T_2	63	42	55	62	50	
123	14	23	12	7	13	
122	16	28	13	6	9	Gruppe 3
118	23	14	7	10	5	
$\bar{X}_3$	17.7	21.7	10.7	7.7	9.0	
T_3	40	41	60	42	51	
$\bar{X}$	27.2	30.2	15.0	11.1	8.7	
s	9.4	9.6	4.3	4.4	2.8	
r_{stab}	.88	.91	.81	.73	.65	

Gruppe					
1 – 2	*	*		*	
1 – 3		*			
2 – 3	*			*	

KL 10

Schüler Nr.	I	II	III	IV	V	
144	13	29	9	4	11	
143	10	26	8	7	11	
142	14	25	10	5	11	
137	14	33	9	11	6	Gruppe 1
136	16	26	8	4	11	
132	14	23	8	7	12	
$\bar{X}_1$	13.5	27.0	8.7	6.3	10.3	
T_1	41	51	44	43	52	
141	32	24	15	9	11	
140	21	23	18	8	11	
139	14	38	13	5	10	Gruppe 2
138	18	48	23	13	3	
134	24	32	20	8	8	
133	18	28	13	8	11	
$\bar{X}_2$	21.2	32.2	17.0	8.5	9.0	
T_2	54	57	60	48	46	
135	26	12	10	16	13	
131	27	13	10	16	9	Gruppe 3
130	20	9	5	17	12	
$\bar{X}_3$	24.3	11.3	8.3	16.3	11.3	
T_3	59	35	43	66	56	
$\bar{X}$	18.7	25.9	11.9	9.2	9.9	
s	6.0	9.7	4.9	4.3	2.5	
r_{stab}	.89	.88	.90	.80	.03	

Gruppe					
1 - 2	*		*		
1 - 3	*	*		*	
2 - 3		*	*	*	

kl. 11

Schüler Nr.	I	II	III	IV	V	
168	9	26	4	6	4	
166	22	41	15	5	5	
165	23	36	9	7	5	
163	15	35	8	6	5	
162	26	28	12	7	5	
161	24	40	12	7	4	
160	12	31	7	5	5	Gruppe 1
154	17	24	8	6	4	
153	18	26	8	8	5	
151	20	26	8	10	6	
150	16	28	6	7	5	
149	24	36	12	8	5	
148	15	30	8	6	6	
147	23	36	10	9	5	
$\bar{X}_1$	18.9	31.6	9.1	6.9	4.9	
T_1	46	55	50	43	50	
167	39	28	23	15	4	Gruppe 2
164	39	39	18	8	4	
$\bar{X}_2$	39.0	33.5	20.5	11.5	4.0	
T_2	71	57	75	56	41	
159	35	9	10	14	6	
158	13	14	4	11	7	
157	18	31	5	14	3	
156	20	11	4	11	4	Gruppe 3
155	34	18	8	13	6	
152	27	12	10	12	6	
146	20	16	8	17	3	
145	18	9	4	11	6	
$\bar{X}_3$	23.1	15.0	6.6	12.9	5.1	
T_3	51	39	44	61	52	
$\bar{\bar{X}}$	22.0	26.3	9.3	9.3	4.9	
s	7.9	9.9	4.5	3.4	1.0	
r_{stab}	.92	.91	.95	.73	.56	

Gruppe					
1 - 2	*		*	*	
1 - 3		*		*	
2 - 3	*	*	*		

Kl. 12

Schüler Nr.	I	II	III	IV	V	
181	15	27	10	6	9	
174	10	21	11	8	11	
172	9	28	8	6	8	Gruppe 1
171	16	32	12	5	7	
170	8	33	8	4	5	
$\bar{X}_1$	11.6	28.2	9.8	5.8	8.0	
T_1	40	50	42	40	51	
180	31	36	23	12	7	
179	19	30	40	11	7	
177	26	28	14	10	7	
176	26	21	9	11	8	Gruppe 2
175	20	38	18	8	5	
173	40	30	16	13	9	
$\bar{X}_2$	27.0	30.5	20.0	10.8	7.2	
T_2	53	54	67	52	47	
178	44	27	11	18	6	Gruppe 3
169	44	19	15	19	12	
$\bar{X}_3$	44.0	23.0	13.0	18.5	9.0	
T_3	66	40	50	69	56	
$\bar{X}$	23.7	28.5	13.0	10.1	7.8	
s	12.3	5.5	4.2	4.5	2.0	
r_{stab}	.96	.52	.86	.92	.52	

Gruppe					
1 - 2	*			*	
1 - 3	*			*	
2 - 3	*			*	

Kl. 13

Schüler Nr.	I	II	III	IV	V	
197	28	9	13	16	11	
193	37	9	14	18	8	
192	41	29	8	17	4	
191	34	24	16	8	6	
190	17	19	7	15	5	Gruppe 1
188	26	28	17	7	9	
184	33	20	20	11	10	
183	18	15	13	7	13	
$\bar{X}_1$	29.3	19.1	13.5	12.4	8.3	
T_1	51	42	48	55	52	
196	22	42	15	5	8	
195	9	35	7	6	8	
187	22	35	12	10	5	Gruppe 2
186	11	33	7	4	8	
185	30	43	23	6	9	
182	11	29	11	4	10	
$\bar{X}_2$	17.5	36.2	12.5	5.8	8.0	
T_2	42	59	47	41	51	
194	54	35	23	16	4	Gruppe 3
189	50	31	22	12	3	
$\bar{X}_3$	52	33	22.5	14.0	3.5	
T_3	69	53	65	58	35	
$\bar{\bar{X}}$	27.7	27.3	14.3	10.1	7.6	
s	13.0	10.1	5.4	4.8	2.8	
r_{stab}	.93	.89	.88	.77	.22	

Gruppe			
1 - 2		*	*
1 - 3	*		
2 - 3	*		

Kl. 14

Schüler Nr.	I	II	III	IV	V	
210	29	46	19	9	9	
206	18	32	15	7	12	
204	14	40	14	4	8	Gruppe 1
203	9	39	12	5	8	
199	13	41	22	8	8	
$\bar{X}_1$	16.6	39.6	16.4	6.6	9.0	
T_1	42	62	57	43	43	
209	22	16	11	9	10	
208	28	13	11	10	14	
205	21	13	8	7	14	Gruppe 2
202	16	13	11	6	14	
$\bar{X}_2$	21.8	13.8	10.3	8.0	13.0	
T_2	47	41	41	46	59	
207	37	21	18	19	10	
201	47	16	9	17	13	
200	40	13	16	12	12	Gruppe 3
198	42	16	12	22	7	
$\bar{X}_3$	41.5	16.5	13.8	17.5	10.5	
T_3	63	43	50	63	49	
$\bar{X}$	25.8	24.5	13.7	10.4	10.7	
s	11.9	12.4	4.0	5.4	2.5	
r_{stab}	.96	.95	.82	.89	.33	

Gruppe					
1 - 2		*			
1 - 3	*	*		*	
2 - 3	*			*	

Kl. 15

Schüler Nr.	I	II	III	IV	V	
228	17	29	10	9	11	
222	13	25	16	9	9	
221	9	21	8	5	9	
217	26	25	26	9	6	Gruppe 1
213	27	14	15	12	5	
211	20	30	11	9	10	
$\bar{X}_1$	18.7	24.0	14.3	8.8	8.3	
T_1	40	45	45	39	52	
227	29	47	23	11	3	
224	31	38	24	18	5	
223	28	34	20	14	10	
220	22	37	26	20	3	Gruppe 2
216	34	27	20	17	7	
215	33	43	21	11	8	
212	24	39	16	13	6	
$\bar{X}_2$	28.7	37.9	21.4	14.9	6.0	
T_2	50	60	57	54	43	
226	48	18	24	16	10	
225	32	18	12	17	9	
219	42	18	18	13	13	
218	40	27	12	21	8	Gruppe 3
214	44	22	11	17	8	
$\bar{X}_3$	41.2	20.6	15.4	16.8	9.6	
T_3	62	41	46	58	57	
$\bar{\bar{X}}$	28.8	28.4	17.4	13.4	7.8	
s	10.3	9.2	5.7	4.3	2.6	
r_{stab}	.89	.88	.86	.64	.44	

Gruppe				
1 - 2	*	*	*	
1 - 3	*		*	
2 - 3	*	*		

Kl. 16

Schüler Nr.	I	II	III	IV	V	
254	36	13	11	13	11	
253	30	27	13	15	5	
252	29	21	12	17	8	
247	34	21	12	14	8	Gruppe 1
237	34	27	12	18	6	
235	42	25	11	20	5	
$\bar{X}_1$	34.2	22.3	11.8	16.2	7.2	
T_1	59	39	39	62	48	
251	20	32	7	11	7	
250	17	33	13	10	7	
249	29	22	16	12	11	
244	34	27	19	14	10	
242	32	23	18	13	9	
239	19	32	14	11	8	Gruppe 2
238	25	29	17	14	6	
236	26	29	18	13	8	
233	24	25	13	12	10	
232	26	28	17	14	10	
231	27	33	18	11	7	
229	23	29	13	11	8	
$\bar{X}_2$	25.2	28.5	15.3	12.2	8.4	
T_2	44	49	49	45	55	
248	37	31	20	13	7	
246	31	32	19	11	8	
245	27	35	21	14	6	
243	31	39	18	13	6	Gruppe 3
241	24	39	19	12	7	
240	24	42	20	12	4	
234	37	34	19	14	7	
230	33	34	20	16	6	
$\bar{X}_3$	30.5	35.8	19.5	13.1	6.4	
T_3	53	60	60	49	44	
$\bar{X}$	28.9	29.3	15.8	13.4	7.5	
s	6.0	6.3	3.6	2.3	1.8	
r_{stab}	.88	.90	.80	.61	.64	

Gruppe					
1 - 2	*	*	*	*	
1 - 3		*	*	*	
2 - 3		*	*		*

Kl. 17

Schüler Nr.	I	II	III	IV	V	
280	14	30	9	8	8	
279	14	30	9	8	8	
277	22	29	10	10	10	
276	27	41	14	12	8	
275	23	35	13	10	9	
274	12	30	10	10	11	
273	21	34	11	10	9	
272	21	29	10	11	8	Gruppe 1
269	28	34	12	11	6	
268	15	26	10	11	7	
267	20	29	11	11	9	
265	21	26	10	9	9	
264	19	34	10	9	8	
261	24	29	12	9	10	
$\bar{X}_1$	20.1	31.1	10.8	9.9	8.6	
T_1	47	57	53	50	43	
278	15	22	8	9	12	
271	31	16	10	13	11	
263	18	21	9	10	11	
262	27	22	12	12	10	
260	21	16	9	10	11	
259	28	14	8	11	12	Gruppe 2
258	20	24	11	9	10	
257	21	22	9	10	11	
256	27	20	12	11	11	
255	26	21	12	11	11	
$\bar{X}_2$	23.4	19.8	10.0	10.6	11.0	
T_2	54	40	47	54	59	
266	23	22	10	4	10	Gruppe 3
$\bar{X}_3$	23.0	22.0	10.0	4.0	10.0	
T_3	53	44	47	15	53	
$\bar{\bar{X}}$	21.5	26.2	10.4	9.9	9.6	
s	4.9	6.6	1.5	1.7	1.5	
r_{stab}	.84	.88	.54	.53	.43	

Gruppe				
1 – 2		*		*
1 – 3	*		*	
2 – 3			*	

Kl. 18

Schüler Nr.	I	II	III	IV	V	
334	10	21	4	8	8	
333	12	33	12	4	9	
330	12	22	8	8	12	
329	21	18	8	12	12	
326	13	19	8	14	6	
323	8	29	5	10	4	
317	15	23	11	13	12	
316	14	16	7	12	12	Gruppe 1
315	17	23	13	13	6	
314	16	18	8	16	12	
313	12	23	7	11	11	
312	18	18	7	12	13	
311	23	16	7	13	12	
$\bar{X}_1$	14.7	21.5	8.1	11.2	9.9	
T_1	42	47	43	47	53	
332	22	45	21	9	6	
324	28	41	26	14	11	Gruppe 2
319	19	47	13	11	2	
$\bar{X}_2$	23.0	44.3	20.0	11.3	6.3	
T_2	50	72	67	47	42	
331	38	22	12	19	6	
328	36	17	11	17	14	
327	33	14	11	16	8	
325	38	20	16	16	7	Gruppe 3
322	36	26	15	16	6	
321	28	14	12	16	12	
320	42	33	16	16	6	
318	40	24	16	16	7	
$\bar{X}_3$	36.4	21.3	13.6	16.5	8.3	
T_3	63	47	54	57	48	
$\bar{X}$	22.9	24.2	11.4	13.0	8.9	
s	10.6	9.1	5.0	5.4	3.2	
r_{stab}	.93	.86	.87	.73	.69	

Gruppe				
1 - 2	*	*	*	
1 - 3	*		*	*
2 - 3	*	*	*	*

Kl.19

Schüler Nr.	I	II	III	IV	V	
363	36	20	16	14	11	
362	37	22	14	12	8	
360	35	21	12	14	8	
357	32	16	13	13	6	
350	33	22	15	12	12	Gruppe 1
345	31	21	14	12	10	
343	34	23	16	12	12	
342	33	21	12	12	10	
336	39	13	11	13	14	
335	39	27	14	12	6	
$\bar{X}_1$	34.9	20.6	13.7	12.6	9.7	
T_1	57	41	48	57	58	
361	41	29	22	13	9	
351	38	26	19	11	9	Gruppe 2
349	35	32	23	11	8	
$\bar{X}_2$	38.0	29.0	21.3	11.7	8.7	
T_2	62	55	72	52	54	
359	25	22	12	8	7	
358	19	33	12	7	9	
356	30	28	14	12	6	
355	23	26	13	12	7	
354	28	25	15	10	6	
353	30	30	17	10	6	
352	28	26	12	12	6	
348	32	26	10	10	6	Gruppe 3
347	30	25	10	13	6	
346	19	39	11	13	5	
344	32	26	15	11	6	
341	30	25	14	9	10	
340	31	32	16	8	8	
339	22	40	10	8	5	
338	34	30	18	11	8	
337	24	31	12	12	7	
$\bar{X}_3$	27.3	29.0	13.2	10.4	6.8	
T_3	44	55	47	45	45	
$\bar{X}$	31.0	26.1	14.2	11.3	7.9	
s	5.7	5.9	3.2	1.8	2.3	
r_{stab}	.84	.84	.84	.22	.81	

Gruppe					
1 - 2		*	*		
1 - 3	*	*		*	*
2 - 3	*		*		

Kl. 20

Schüler Nr.	I	II	III	IV	V	
385	25	42	15	14	4	
384	27	30	16	8	8	
376	34	40	15	8	6	
375	28	41	15	8	7	
374	21	40	16	8	7	
373	28	36	15	9	7	Gruppe 1
371	23	34	14	8	9	
369	19	33	10	11	8	
368	24	40	14	12	6	
366	33	37	17	10	7	
365	22	42	17	7	5	
$\bar{X}_1$	25.8	37.7	14.9	9.4	6.7	
T_1	47	58	55	46	42	
383	30	14	8	9	11	
382	30	13	8	15	12	
379	23	15	12	17	13	
377	26	20	11	8	12	
372	25	26	11	10	12	Gruppe 2
370	22	25	11	9	12	
367	32	19	13	13	12	
364	24	20	12	9	12	
$\bar{X}_2$	26.5	19.0	10.8	11.3	12.0	
T_2	48	39	42	52	62	
381	35	38	21	14	9	
380	32	28	11	14	9	Gruppe 3
378	39	21	12	13	8	
$\bar{X}_3$	35.3	29.0	14.7	13.7	8.7	
T_3	66	49	54	61	49	
$\bar{X}$	27.4	29.7	13.4	10.6	8.9	
s	5.1	9.8	3.1	2.8	2.6	
r_{stab}	.73	.94	.68	.74	.91	

Gruppe				
1 - 2		*	*	*
1 - 3	*	*		*
2 - 3	*	*		*

Kl. 21

Schüler Nr.	I	II	III	IV	V	
421	23	23	12	10	8	
420	32	37	17	11	7	
419	31	31	16	26	3	
417	19	44	17	12	5	
413	23	31	12	11	10	
412	22	26	13	11	9	
411	38	40	15	12	6	
410	18	24	12	11	11	
409	34	24	12	16	9	
408	35	35	17	13	8	
407	23	30	9	13	11	Gruppe 1
406	28	37	20	10	7	
404	24	36	15	12	6	
402	20	25	14	12	8	
400	25	27	15	11	9	
394	32	33	14	16	6	
392	18	15	6	11	12	
391	23	18	10	16	10	
389	32	16	11	20	7	
388	22	21	9	12	11	
$\bar{X}_1$	26.1	28.7	13.3	13.3	8.2	
T_1	51	48	51	52	53	
418	16	33	12	5	7	
416	16	34	14	7	7	
415	13	46	9	7	6	
414	12	42	9	8	7	
405	21	39	9	7	5	
403	17	33	9	8	5	
401	19	29	13	8	8	
396	18	46	14	9	4	Gruppe 2
393	15	29	12	8	7	
390	16	39	11	8	5	
387	16	34	13	7	6	
386	12	29	8	7	5	
$\bar{X}_2$	15.9	36.1	11.1	7.4	6.0	
T_2	41	57	45	40	43	
399	47	18	14	17	12	
398	51	23	21	19	9	
397	48	36	23	16	7	
395	47	21	14	19	11	Gruppe 3
$\bar{X}_3$	48.3	24.5	18.0	17.8	9.8	
T_3	72	42	64	63	60	
$\bar{\bar{X}}$	25.2	30.7	13.0	11.8	7.6	
s	10.5	8.2	3.6	4.5	2.3	
r_{stab}	.92	.87	.73	.79	.43	
Gruppe						
1 - 2	*	*		*	*	
1 - 3	*		*	*		
2 - 3	*	*	*	*	*	

Kl. 22

Schüler Nr.	I	II	III	IV	V	
457	42	32	23	12	7	
455	42	31	24	10	9	
452	41	34	24	9	4	Gruppe 1
440	31	40	19	10	6	
438	40	32	21	11	9	
428	35	37	17	10	7	
$\bar{X}_1$	38.5	34.3	21.3	10.3	7.0	
T_1	59	60	64	48	50	
456	15	20	9	8	6	
454	23	26	12	9	9	
453	14	32	9	7	8	
451	9	37	8	5	3	
447	15	28	6	8	6	
446	25	28	11	11	4	
444	22	39	10	9	4	
442	20	16	12	10	10	
441	15	25	11	8	7	Gruppe 2
433	17	18	10	9	8	
432	12	23	13	6	11	
431	24	44	14	8	6	
429	23	18	12	10	7	
427	28	39	14	10	7	
426	26	36	12	7	5	
424	27	31	13	10	7	
422	17	13	7	8	11	
$\bar{X}_2$	19.5	27.8	10.7	8.4	7.0	
T_2	41	52	43	43	50	
450	30	20	12	13	5	
449	27	17	12	12	6	
448	39	19	16	19	6	
445	36	23	10	13	4	
443	42	25	24	12	10	
439	28	13	14	14	9	
437	39	20	16	18	8	
436	38	17	16	13	10	Gruppe 3
435	33	17	11	11	8	
434	31	20	10	12	5	
430	42	15	11	18	4	
425	48	21	24	14	9	
423	43	20	21	16	8	
$\bar{X}_3$	36.6	19.0	15.2	14.2	7.1	
T_3	57	42	52	60	50	
$\bar{X}$	28.9	25.7	14.1	10.8	7.0	
s	10.5	8.5	5.1	3.3	2.1	
r_{stab}	.94	.87	.91	.75	.14	

Gruppe				
1 - 2	*		*	
1 - 3		*	*	*
2 - 3	*	*	*	*

Kl. 23

Schüler Nr.	I	II	III	IV	V	
558	31	32	11	11	8	
557	34	39	16	11	5	
555	18	35	13	8	3	
547	22	38	14	8	7	Gruppe 1
542	19	35	15	9	9	
540	20	40	20	10	8	
538	24	33	13	12	7	
$\bar{X}_1$	24.0	36.0	14.6	9.9	6.7	
T_1	43	60	45	43	46	
556	36	30	21	10	8	
554	50	27	24	23	4	
553	50	28	25	20	2	
552	56	31	28	26	6	
551	46	36	22	17	8	
550	44	21	18	18	8	Gruppe 2
549	48	31	25	19	10	
548	38	34	20	22	9	
546	45	30	28	17	6	
545	38	24	21	16	11	
539	37	22	16	11	12	
$\bar{X}_2$	44.4	28.5	22.5	17.2	7.5	
T_2	60	50	59	57	49	
544	29	27	13	12	9	
543	27	20	16	10	11	
541	26	12	11	8	8	
537	16	21	10	8	11	Gruppe 3
536	23	29	13	8	10	
535	19	16	10	16	6	
534	17	17	9	7	10	
$\bar{X}_3$	22.4	20.3	11.7	9.9	9.3	
T_3	41	39	40	43	56	
$\bar{X}$	32.5	28.3	17.3	13.5	7.8	
s	11.9	7.4	5.7	5.4	2.5	
r_{stab}	.94	.81	.92	.82	.42	

Gruppe					
1 - 2	*	*	*	*	
1 - 3		*			
2 - 3	*	*	*	*	

Kl. 25

Schüler Nr.	I	II	III	IV	V	
579	33	28	18	12	8	
574	41	18	21	17	12	
573	43	35	20	10	7	Gruppe 1
572	41	18	10	9	11	
571	43	28	17	9	4	
569	37	39	22	6	4	
$\bar{X}_1$	39.7	27.7	18.0	10.5	9.5	
T_1	55	54	62	59	53	
578	25	14	11	8	14	
576	31	20	8	10	4	
568	25	14	11	8	14	Gruppe 2
564	16	17	12	5	8	
563	24	11	11	8	13	
561	18	16	8	5	8	
559	26	14	8	4	13	
$\bar{X}_2$	23.6	15.1	9.9	6.9	10.6	
T_2	50	32	44	48	56	
577	18	25	8	9	9	
575	12	35	13	5	8	
570	11	31	11	6	8	
567	20	32	16	11	9	Gruppe 3
566	8	26	8	5	9	
565	10	30	10	4	4	
562	10	30	12	5	9	
560	13	26	7	4	8	
$\bar{X}_3$	12.8	29.4	10.6	6.1	8.0	
T_3	40	57	46	45	48	
$\bar{X}$	24.2	24.2	12.5	7.6	8.7	
s	11.6	7.9	4.5	3.2	3.0	
r_{stab}	.93	.86	.89	.80	.66	

Gruppe				
1 - 2	*	*	*	
1 - 3	*		*	*
2 - 3	*	*		

Kl. 26

Schüler Nr.	I	II	III	IV	V	
602	15	29	11	6	8	
596	23	26	14	9	11	
595	11	29	8	6	9	
589	18	30	12	5	8	
588	12	20	11	7	10	
587	15	38	13	6	8	Gruppe 1
585	19	33	12	8	5	
584	22	43	21	6	5	
582	16	43	10	9	4	
581	19	22	14	7	12	
580	24	29	15	8	12	
$\bar{X}_1$	17.6	31.1	12.8	7.0	8.4	
T_1	41	55	46	42	48	
601	42	41	26	11	7	
600	45	25	20	16	7	
598	42	33	15	14	4	Gruppe 2
592	33	28	23	8	10	
591	48	33	18	11	4	
$\bar{X}_2$	42.0	32.0	20.4	12.0	6.4	
T_2	62	56	59	54	42	
599	29	13	9	11	12	
597	36	19	17	8	10	
594	37	18	24	18	14	
593	36	12	7	17	10	Gruppe 3
590	34	16	21	13	10	
586	50	18	13	17	12	
583	21	17	10	13	13	
$\bar{X}_3$	34.7	16.1	14.4	13.9	11.6	
T_3	56	38	49	59	59	
$\bar{X}$	28.1	26.7	15.0	10.2	8.9	
s	11.9	9.1	5.3	4.0	3.0	
r_{stab}	.93	.89	.89	.74	.85	

Gruppe					
1 - 2	*		*	*	
1 - 3	*	*		*	*
2 - 3		*			*

Kl. 26

Schüler Nr.	I	II	III	IV	V	
624	9	26	7	4	9	
623	14	24	13	7	6	
622	8	22	5	4	9	
621	14	28	13	5	8	
618	10	33	13	4	7	Gruppe 1
617	16	30	10	5	8	
614	13	24	11	4	9	
611	10	21	9	9	5	
$\bar{X}_1$	11.8	26.0	10.1	5.3	7.6	
T_1	41	44	44	40	55	
620	22	28	12	12	6	
619	30	27	13	10	10	
616	20	32	14	7	4	
615	22	29	12	13	4	
613	24	45	19	7	3	
612	15	31	13	11	2	
610	17	34	13	10	4	Gruppe 2
609	20	31	13	11	4	
608	25	36	15	8	6	
607	30	23	6	10	8	
606	26	32	14	14	5	
604	22	28	12	13	10	
603	37	31	13	15	7	
$\bar{X}_2$	23.8	31.3	13.0	10.8	5.6	
T_2	53	54	53	54	46	
605	51	24	13	18	7	Gruppe 3
$\bar{X}_3$	51	24	13	18	7	
T_3	81	41	53	73	53	
$\bar{\bar{X}}$	20.7	29.0	12.0	9.1	6.4	
s	9.9	5.3	3.0	3.9	2.3	
r_{stab}	.97	.62	.85	.84	-.06	

Gruppe			
1 - 2		*	
1 - 3	*	*	*
2 - 3	*	*	*

Kl. 27

Schüler Nr.	I	II	III	IV	V	
646	19	35	16	8	7	
644	10	34	14	8	9	
641	32	42	16	10	8	
637	9	40	12	9	9	
634	29	35	18	13	8	Gruppe 1
632	22	38	10	12	7	
631	20	37	16	9	9	
627	19	33	7	8	8	
625	24	41	21	10	8	
$\bar{X}_1$	20.4	37.2	14.4	9.7	8.1	
T_1	43	60	48	46	44	
645	21	20	7	9	10	
642	24	21	15	9	9	
639	32	18	14	13	14	
638	36	24	17	12	9	
636	29	27	13	10	9	Gruppe 2
635	19	19	10	7	12	
629	24	29	10	14	5	
628	35	16	13	10	12	
626	34	26	17	14	9	
$\bar{X}_2$	28.2	22.2	12.9	10.9	9.9	
T_2	50	40	45	52	52	
643	40	31	23	12	9	
640	41	27	21	13	12	
633	44	34	23	13	10	Gruppe 3
630	46	26	21	8	13	
$\bar{X}_3$	42.8	29.5	22.0	11.5	11.0	
T_3	66	50	64	46	58	
$\bar{X}$	27.7	29.7	15.2	10.5	9.4	
s	10.0	7.7	4.7	2.2	2.1	
r_{stab}	.92	.87	.84	.68	.75	

Gruppe					
1 - 2	*			*	
1 - 3	*			*	
2 - 3	*			*	

Kl. 28

Schüler Nr.	I	II	III	IV	V	
683	31	17	5	16	8	
681	25	19	16	15	10	
678	20	22	7	12	10	
675	41	12	6	11	11	
674	38	11	8	13	7	
672	27	19	8	10	11	
671	32	13	7	13	8	
669	22	22	10	10	8	Gruppe 1
667	41	26	12	9	5	
666	20	18	10	8	12	
664	16	21	11	8	11	
660	40	24	7	15	4	
659	15	19	11	9	10	
658	25	19	8	9	10	
657	18	24	5	8	11	
649	28	17	11	11	11	
$\bar{x}_1$	27.4	18.9	8.9	11.0	9.2	
T_1	53	40	43	54	57	
682	36	49	16	7	2	
677	50	40	24	8	6	Gruppe 2
670	44	40	24	9	4	
$\bar{x}_2$	43.3	43.0	21.3	8.0	4.0	
T_2	69	66	67	43	36	
680	22	30	12	11	8	
679	14	41	20	8	4	
676	16	43	8	9	5	
673	26	35	21	15	8	
668	23	33	22	13	6	
665	17	28	12	8	5	
663	10	38	12	6	6	
662	13	32	14	8	8	
661	13	40	9	10	3	Gruppe 3
656	19	31	11	9	8	
655	27	37	19	10	6	
654	10	31	10	5	8	
653	18	25	19	11	10	
652	21	32	11	9	6	
651	22	29	17	9	6	
650	15	28	12	8	6	
648	20	32	12	9	8	
647	15	28	10	8	6	
$\bar{x}_3$	17.8	32.9	13.9	9.2	6.5	
T_3	43	56	53	47	46	
$\bar{x}$	24.1	27.7	12.4	9.9	7.5	
s	10.0	9.3	5.2	2.6	2.5	
r_{stab}	.92	.91	.95	.63	.61	
Gruppe						
1 - 2	*	*	*			
1 - 3	*	*	*		*	
2 - 3	*	*	*		*	

Kl. 29

Schüler Nr.	I	II	III	IV	V	
708	25	39	16	8	10	
706	27	43	21	10	10	
705	39	31	20	12	9	
702	13	32	8	11	8	
700	12	32	12	6	7	
698	21	36	10	5	5	Gruppe 1
692	18	39	16	9	8	
690	17	30	12	5	10	
688	18	31	15	9	8	
687	32	34	15	12	10	
685	12	33	18	8	9	
684	17	27	13	7	10	
$\bar{X}_1$	20.9	33.9	14.7	8.5	8.7	
T_1	48	56	54	42	51	
707	43	15	10	20	13	
704	42	18	4	13	11	
703	39	25	17	23	8	
695	19	13	17	20	13	
693	18	19	14	10	12	Gruppe 2
691	29	14	5	23	9	
689	34	18	17	12	8	
686	22	15	9	12	13	
$\bar{X}_2$	30.8	17.1	11.6	16.6	10.9	
T_2	58	37	47	55	59	
701	16	40	9	17	4	
699	14	35	9	24	4	
697	10	35	13	17	4	
696	25	27	11	20	5	Gruppe 3
694	16	40	12	21	4	
$\bar{X}_3$	16.2	35.4	10.8	19.8	4.2	
T_3	43	57	45	61	35	
$\bar{X}$	23.1	28.8	12.9	13.4	8.5	
s	9.8	9.1	4.3	5.9	2.8	
r_{stab}	.90	.82	.65	.82	.66	
Gruppe						
1 - 2	*	*		*	*	
1 - 3				*	*	
2 - 3	*	*			*	

Kl. 30

Schüler Nr.	I	II	III	IV	V	
745	20	39	14	8	6	
742	21	32	13	9	6	
739	27	37	16	11	6	
737	24	34	14	7	6	
736	19	44	13	8	4	
732	20	37	9	10	8	
731	21	40	13	8	8	
730	18	35	8	8	6	
729	20	28	11	9	7	Gruppe 1
728	23	36	13	9	9	
727	14	31	7	8	8	
726	12	32	6	8	8	
723	24	42	13	10	5	
720	16	31	10	9	7	
715	20	35	12	8	7	
712	24	36	13	8	7	
713	15	29	8	9	9	
709	16	35	8	8	5	
$\bar{X}_1$	19.7	35.2	11.2	8.6	6.8	
T_1	43	58	49	44	50	
744	22	25	13	8	7	
743	38	23	18	12	8	
740	28	24	14	12	6	
735	35	24	12	17	7	
734	31	28	17	11	8	
733	33	25	16	11	7	
725	37	30	14	13	5	Gruppe 2
723	24	42	13	10	5	
719	36	22	14	12	5	
718	31	27	8	8	6	
717	36	18	15	15	9	
714	30	26	15	12	5	
711	34	24	12	9	7	
710	34	29	18	10	6	
$\bar{X}_2$	32.1	14.0	14.2	11.4	6.5	
T_2	58	31	52	54	48	
741	21	12	6	14	4	
738	22	18	11	12	9	
724	27	18	9	9	8	Gruppe 3
721	15	21	8	8	6	
716	29	14	11	9	12	
$\bar{X}_3$	22.8	16.6	9.0	10.4	7.8	
T_3	47	34	46	51	56	
$\bar{X}$	25.4	28.7	12.2	10.2	6.8	
s	7.9	7.7	3.4	2.8	1.7	
r_{stab}	.93	.86	.88	.77	.34	
Gruppe						
1 - 2	*	*	*	*		
1 - 3		*				
2 - 3	*	*	*			

Kl.31

Schüler Nr.	I	II	III	IV	V	
776	25	28	15	14	8	
775	24	25	10	13	7	
773	27	34	17	13	6	
772	24	28	13	15	6	
769	25	34	17	14	6	
766	25	25	13	13	9	
763	12	30	10	13	7	
762	23	38	16	13	6	Gruppe 1
761	25	29	12	13	9	
760	27	27	12	13	9	
759	25	23	14	16	10	
757	18	38	16	11	6	
756	28	30	17	13	8	
755	24	33	14	13	7	
753	20	31	12	12	5	
747	25	35	15	13	7	
$\bar{X}_1$	23.6	30.5	13.9	13.3	7.3	
T_1	42	54	47	47	50	
774	38	20	17	17	6	
771	26	27	12	15	6	
770	35	23	13	17	8	
768	34	21	12	18	7	
767	40	13	13	20	11	
765	32	24	16	13	9	Gruppe 2
764	35	28	20	15	8	
758	31	23	13	19	6	
749	27	18	12	15	9	
$\bar{X}_2$	33.1	21.9	14.2	16.6	7.8	
T_2	55	39	48	59	53	
754	39	32	18	17	6	
752	42	24	18	16	7	
751	37	32	16	18	5	
750	34	30	16	16	6	
748	40	35	17	14	6	Gruppe 3
746	42	30	16	14	9	
$\bar{X}_3$	39.0	30.5	16.8	15.8	6.5	
T_3	63	54	59	55	45	
$\bar{X}$	29.3	28.0	14.6	14.7	7.3	
s	7.4	5.7	2.5	2.2	1.5	
r_{stab}	.94	.90	.81	.69	.42	
Gruppe						
1 - 2	*	*		*		
1 - 3	*		*	*		
2 - 3	*	*				

Kl. 32

Schüler Nr.	I	II	III	IV	V	
797	17	22	8	13	10	
787	23	33	13	13	8	
786	26	32	14	11	7	
785	17	22	11	11	7	
784	19	22	10	11	7	Gruppe 1
783	19	20	11	11	8	
782	18	18	8	12	8	
781	10	29	12	9	9	
779	13	23	5	9	8	
778	19	23	10	10	10	
777	21	33	13	10	9	
$\bar{X}_1$	18.4	25.2	10.5	10.8	8.3	
T_1	43	51	48	42	52	
796	37	26	6	19	6	
795	34	36	21	16	7	
793	34	31	18	13	8	Gruppe 2
789	33	37	16	16	7	
780	37	27	21	16	6	
$\bar{X}_2$	35.0	31.4	16.4	16.0	6.8	
T_2	65	59	61	56	43	
794	20	17	6	17	6	
792	24	11	7	23	8	
791	27	27	15	16	6	Gruppe 3
790	27	16	8	15	7	
788	18	14	11	18	12	
$\bar{X}_3$	23.2	17.0	9.4	17.8	7.8	
T_3	50	39	45	61	50	
$\bar{X}$	23.5	24.7	11.6	13.8	7.8	
s	7.6	7.1	4.5	3.6	1.5	
r_{stab}	.89	.81	.95	.59	.21	

Gruppe				
1 - 2	*		*	*
1 - 3		*		*
2 - 3	*	*	*	

Kl. 33

Schüler Nr.	I	II	III	IV	V	
832	42	13	11	17	11	
830	25	15	8	12	10	
829	24	23	12	12	8	
827	34	22	13	12	6	
824	27	16	11	14	10	
820	26	15	12	12	6	
819	28	16	8	14	11	
818	30	17	11	14	10	
817	26	20	11	14	11	
816	24	22	15	10	8	
815	26	15	14	14	11	
814	38	19	18	11	7	
812	36	23	15	14	8	Gruppe 1
811	28	29	15	12	2	
810	30	13	12	17	11	
808	26	21	19	14	10	
806	28	25	20	12	9	
804	35	22	16	15	9	
803	39	17	9	14	8	
801	31	18	10	17	10	
800	31	25	13	14	8	
798	22	14	8	11	3	
$\bar{X}_1$	29.8	19.1	12.8	13.8	8.5	
T_1	52	44	48	55	52	
831	33	36	19	14	5	
826	33	32	20	12	5	
823	27	42	18	11	5	
822	34	39	18	13	6	Gruppe 2
813	42	37	14	13	3	
807	46	39	19	13	9	
$\bar{X}_2$	35.8	37.5	18.0	12.7	5.5	
T_2	60	65	63	50	40	
828	20	34	10	11	6	
825	22	33	16	11	10	
821	18	29	12	9	7	
809	12	25	8	14	5	
805	16	21	12	8	11	Gruppe 3
802	22	39	14	11	9	
799	15	33	10	9	8	
$\bar{X}_3$	17.9	30.6	11.7	10.4	8.0	
T_3	36	57	45	39	50	
$\bar{X}$	28.5	24.5	13.5	12.7	7.9	
s	7.7	8.6	3.5	2.1	2.5	
r_{stab}	.88	.90	.90	.47	.65	
Gruppe						
1 - 2		*	*		*	
1 - 3	*	*		*		
2 - 3	*	*	*			

Kl. 34

Schüler Nr.	I	II	III	IV	V	
880	29	43	23	11	5	
873	41	25	15	12	6	
872	40	26	15	12	6	
870	37	29	18	8	11	
868	23	29	20	10	11	Gruppe 1
866	36	26	20	10	9	
865	23	26	18	10	12	
861	31	28	16	10	5	
860	37	29	19	14	5	
859	39	20	15	12	7	
855	36	27	19	10	12	
$\bar{X}_1$	33.2	28.0	17.1	10.8	8.0	
T_1	53	55	55	48	47	
879	13	26	3	8	8	
877	10	34	9	8	5	
867	18	31	15	8	10	
864	10	37	9	9	7	Gruppe 2
857	14	29	12	9	9	
856	10	28	8	7	8	
854	18	27	9	7	6	
$\bar{X}_2$	13.3	30.3	10.0	8.0	7.6	
T_2	36	58	40	42	45	
878	32	17	8	9	11	
876	50	16	21	21	8	
875	39	9	8	12	13	
874	35	16	21	12	8	
871	44	11	9	9	12	Gruppe 3
869	33	20	14	15	12	
863	36	15	13	20	13	
862	20	18	9	14	11	
858	52	15	19	25	13	
$\bar{X}_3$	37.9	15.2	13.6	15.2	11.2	
T_3	57	38	48	58	58	
$\bar{X}$	29.9	24.3	14.8	11.6	8.9	
s	12.1	7.8	4.8	4.3	2.8	
r_{stab}	.95	.83	.93	.79	.73	

Gruppe					
1 – 2	*		*		
1 – 3		*	*	*	*
2 – 3	*	*		*	*

Kl. 35

Schüler Nr.	I	II	III	IV	V	
910	35	35	20	15	8	
909	27	46	15	10	6	
905	19	38	17	11	9	
904	15	39	17	10	8	
903	27	37	20	10	8	
901	14	31	13	10	6	Gruppe 1
900	17	39	15	10	8	
897	31	33	18	13	8	
896	18	41	17	10	9	
895	12	43	13	8	10	
894	32	29	18	14	8	
887	30	37	19	15	7	
883	26	36	18	10	10	
$\bar{X}_1$	23.3	37.2	16.9	11.2	8.1	
T_1	45	59	52	44	50	
908	40	26	17	16	7	
907	41	28	20	17	7	
906	51	34	14	20	8	
898	43	38	20	15	6	
893	43	17	18	20	5	
889	44	26	19	17	10	Gruppe 2
888	40	29	20	17	9	
885	44	26	22	18	9	
882	41	22	15	18	8	
881	41	23	17	14	9	
$\bar{X}_2$	42.8	26.9	18.2	17.2	7.8	
T_2	62	46	56	61	48	
902	30	23	13	16	9	
899	24	25	11	13	8	
892	25	27	17	9	6	
891	13	14	4	10	9	
890	18	25	15	8	8	Gruppe 3
886	17	22	12	13	10	
884	19	24	13	12	10	
$\bar{X}_3$	20.9	22.9	12.1	11.6	8.6	
T_3	43	40	39	45	54	
$\bar{X}$	29.3	30.4	16.2	13.3	8.1	
s	11.3	7.8	3.6	3.5	1.4	
r_{stab}	.96	.87	.71	.69	– 0.07	
Gruppe						
1 – 2	*	*		*		
1 – 3		*	*			
2 – 3	*		*	*		

Kl. 36

Schüler Nr.	I	II	III	IV	V	
938	10	23	11	6	10	
936	17	34	16	7	9	
935	18	23	4	9	9	
934	15	31	3	6	8	
930	15	25	13	6	9	
929	15	19	11	13	9	
928	16	31	14	5	10	Gruppe 1
925	9	24	7	5	9	
924	49	33	19	19	7	
919	9	36	12	5	4	
917	15	21	11	5	12	
915	12	29	11	6	7	
914	11	32	12	4	9	
912	14	26	11	5	10	
$\bar{X}_1$	16.1	27.6	12.5	7.2	8.7	
T_1	44	45	44	46	52	
937	33	38	24	10	9	
933	28	47	24	9	8	
932	19	42	13	10	4	
927	32	39	16	15	4	
921	35	38	15	9	7	Gruppe 2
920	26	42	17	10	8	
916	32	48	20	7	5	
913	22	39	16	9	8	
911	34	41	23	12	8	
$\bar{X}_2$	29.0	41.6	18.7	10.1	6.8	
T_2	55	62	57	53	42	
931	28	21	16	11	11	
926	38	31	24	16	9	
924	49	33	19	19	7	
923	45	36	20	14	10	Gruppe 3
918	35	22	18	12	12	
$\bar{X}_3$	39	28.6	19.4	13.4	9.8	
T_3	64	46	59	62	57	
$\bar{X}$	23.1	32.0	15.4	8.9	8.4	
s	11.1	8.1	4.5	3.9	2.0	
r_{stab}	.95	.84	.88	.87	.33	
Gruppe						
1 - 2	*	*	*			
1 - 3	*		*	*		
2 - 3		*			*	

L e b e n s l a u f

Ich, Klaus Walter Helmut Tscherner, wurde am 9. März 1939 in Gelsenkirchen als Sohn des Grubenobersteigers Erwin Tscherner, geb. am 3. 11. 1904 und seiner Ehefrau Cläre Tscherner, geborene Becker, geb. am 18. 4. 1907, geboren.

Von Ostern 1946 bis Ostern 1950 besuchte ich die Volksschule in Bottrop. Von Ostern 1950 bis Ostern 1954 ging ich zum Gymnasium in Bottrop. Von Ostern 1954 bis Ostern 1957 war ich Schüler des Gymnasiums in Burgsteinfurt. Während dieser Zeit wohnte ich im dortigen Alumnat. Von Ostern 1957 bis Ostern 1959 besuchte ich das Gymnasium in Dorsten. Danach arbeitete ich vom 4. 5. 1959 bis zum 30. 9. 1959 als Gedinge Schlepper auf der Zeche Matthias Stinnes III/IV. Vom 5. 10. 1959 bis zum 5. 10. 1964 war ich Soldat der deutschen Bundeswehr. Während dieser Zeit, vom Oktober 1961 bis zum 4. 2. 1964, besuchte ich das Abendgymnasium der Stadt und des Landkreises Aachen. Dort wurde mir am 4. 2. 1964 das Zeugnis der Reife zuerkannt. Am 2. 5. 1964 nahm ich an der Phillips-Universität in Marburg das Studium der Psychologie auf.

Das Vordiplom legte ich im Oktober 1967, das Hauptdiplom in Psychologie nach dem WS 1969/70 in Marburg ab.

Danach arbeitete ich in der Sonderpädagogischen Beratungsstelle der PH Ruhr in Dortmund. Vom 15. 2. 1971 bis 28. 2. 1973 war ich wissenschaftlicher Mitarbeiter der "Forschungsgruppe zur wissenschaftlichen Begleitung des Gesamtschulversuchs NW" an der PH Ruhr, Abt. Dortmund. Meine Aufgabenschwerpunkte in dieser Forschungsgruppe lagen in der Beratung empirischer Forschungsvorhaben im curricularen und sozialwissenschaftlichen Bereich. Zudem leitete ich die Durchführung und Auswertung der Eingangserhebungen an allen Gesamtschulen in NW.

Seit dem 1. 3. 1973 bin in Wiss. Assistent (m. d. V. b.) am Seminar für Psychologie und Soziologie der Abteilung Hagen der PH Ruhr.

Klaus Tscherner